

colecția

# BIBLIOTECA DE ȘTIINȚE

seria • Științe umaniste

FILOSOFIA MENTALULUI

Intenționalitate  
și  
experiment



Editura Științifică



Colecția **BIBLIOTECA DE ȘTIINȚE**

libris, Seria **ȘTIINȚE UMANISTE**

1994

# **FILOSOFIA MENTALULUI**

## **INTENȚIONALITATE ȘI EXPERIMENT**

**Coordonator: ANGELA BOTEZ**



**EDITURA ȘTIINȚIFICĂ**  
**BUCUREȘTI, 1996**

**Redactor: ARGENTINA FIRUȚĂ**

**Grafician: SIMONA DUMITRESCU**

**© Editura Științifică, București, 1996**

**Lucrare editată cu sprijinul Ministerului Cercetării și Tehnologiei – Colegiul  
Consultativ pentru Cercetare Științifică și Dezvoltare Tehnologică.**

**ISBN 973-44-0197-1**

# SUMAR

<i>Studiu introductiv</i> – Introducere în filosofia mentalului – <b>Angela Botez</b> (București) .....	7
---------------------------------------------------------------------------------------------------------	---

## PARTEA I-A ORIENTĂRI ȘI CONCEPTE

<b>Ted Honderich</b> (Londra) – Conștiință, funcționalism neural și subiectivitate reală (trad. A. Botez) .....	41
<b>David Papineau</b> (Londra) – Adevăr și teleologie (trad. G. Nagâț) .....	58
<b>Tim Crane</b> (Londra) – Cauzalitatea mentală (trad. L. Staicu) .....	89
<b>John J. Haldane</b> (St. Andrews) – Naturalism și problema intenționalității (trad. M. Vornicu și A. Botez) .....	117
<b>Lars Bergstrom</b> (Stockholm) – Explicația și interpretarea acțiunii (trad. M. Vornicu și A. Botez) .....	140
<b>Graham J. Lock</b> (Oxford) – Filosofia analitică și teoria psihanalitică (trad. A. Botez) .....	162
<b>I. Iubomira Miroș</b> (București) – Conștiință de sine și cunoaștere morală ....	172
<b>Mircea Malița și Mihaela Malița</b> (București) – Problema unității științei și inteligența artificială .....	188

## PARTEA II-A PROBLEMA *MIND-BODY* ȘI EXPERIMENTUL MENTAL

<b>Karl R. Popper</b> (Londra) – Problema minte-trup (trad. L. Staicu) .....	197
<b>Kathleen V. Wilkes</b> (Oxford) – Minte și trup ( <i>Mind-body</i> )(trad. M. Vornicu și A. Botez).....	209
<b>James R. Brown</b> (Toronto) – Experimentele mentale: o ipoteză platonistă (trad. D. Gheorghiu).....	229
<b>Roy A. Sorensen</b> (New-York) – Experimentele mentale și epistemologia legilor (trad. D. Gheorghiu) .....	243
<b>Ion Mânzat</b> (București) – Experimentul mental din perspectiva psihologică.....	274
<b>Gheorghe Ștefan</b> (București) – Coerența mentalului și formalului în experiment .....	293
<b>G. G. Constandache</b> (București) – Mental și experimental în științele cognitive .....	308
<b>Bibliografie selectivă</b> (A. Botez).....	319
<b>Indice de nume</b> (A. Botez) .....	327



# INTRODUCERE ÎN FILOSOFIA MENTALULUI

## STUDIU INTRODUCTIV

Angela BOTEZ

### I. FILOSOFIA MENTALULUI ȘI ȘTIINȚA COGNITIVĂ

În încercarea de a arăta ce este filosofia mentalului<sup>\*</sup>, vom prezenta, în cele ce urmează, câteva definiții, concepte și teme, orientări și personalități ale domeniului. Filosofia mentalului reprezintă abordarea ontologică (a naturii), epistemologică (a explicației și înțelegerii), semantică (a semnificației) a eului uman, a actelor sale cognitive, a conștiinței și comportamentului intențional. Ea se constituie astăzi și ca fundamentare metodologică a științelor cognitive, discipline noi apărute în aria cercetării stărilor și proceselor conștiinței.

Ca ramură a filosofiei ea se delimitează, dar și interacționează cu filosofia limbajului și a științei, cu logica și metafizica. Se raportează, de asemenea, la rezultatele științifice obținute de psihologie, cibernetică, teoria Inteligenței Artificiale, științele cognitive.

**Izvoare ale filosofiei mentalului găsim în istoria filosofiei<sup>1</sup>** încă de la gânditorii antici: în cognitivismul lui Socrate și Platon și în funcționalismul naturalist al lui Aristotel, continuând cu elemente de reprezentationalism mental în scolastica medievală, apoi cu dualismul cartezian, fizicalismul și introspecționismul lui Spinoza, cu concepția subiectivistă a lui Berkeley, care identifică existența cu percepția (*esse est percipi*), dar mai ales cu teoria despre calități și conținut mental, a lui Hume și cu cea despre comunicare și calcul mental, a lui Leibniz.

Mai putem nota un anume tip de funcționalism și cognitivism la Kant, concepția fenomenologică a intenționalității la Husserl și

---

<sup>\*</sup> Am tradus în limba română „Philosophy of Mind“ prin *filosofia mentalului* dar și prin *filosofia minții*. Ca și în cazul noțiunilor: *background*, *supervenience* sau *mind-body*, am folosit, uneori, în timpul lucrării și expresiile englezești, întrucât cuvintele limbii române nu exprimă cu exactitate ceea ce semnifică acești termeni în engleză. Pentru traducerea unor expresii ne-am consultat, în Anglia, cu Tim Crane, Ted Honderich și David Papineau, care sunt de părere că „filosofia mentalului“ poate fi un echivalent acceptabil pentru „mind philosophy“.

Brentano, unitatea logic-psihic, fizic-practic în viziunea pragmatică (Peirce și James) și concepția asupra fenomenelor mentale a școlii analitice.

Există o serie de filosofi pe care-i putem considera chiar precursori ai filosofiei contemporane a mentalului datorită unor concepte și metode pe care le utilizează. Dintre ei amintim pe: Occam cu limbajul mental, Descartes cu dualismul spirit-materie, Reid cu doctrina conștiinței, Husserl cu orientarea reprezentării, Ramsey cu teoria funcționalistă a credinței (*belief*), Wundt cu procese psiho-mentale și părțile lor, Wittgenstein cu „jocurile de limbaj” și „formele de viață.” De asemenea, alții cum ar fi: Brentano, Titchener, Bergson, Jung, Freud, Teilhard de Chardin, Merleau-Ponty, Habermas, Ayer, Chomsky. La câțiva dintre aceștia vom reveni mai jos. Găsim teme ale filosofiei mentalului și la unii filosofi români: C. Rădulescu-Motru, L. Blaga, M. Florian, C. Odobleja etc.

**Conceptele și noțiunile prezente în majoritatea lucrărilor de filosofie psihologică și a mentalului sunt:** 1) Eu, conștiință, intenționalitate; 2) Stări, acte și procese mentale – evenimente mentale; 3) Cunoscător, cunoscut, cunoaștere; 4) Subdeterminare (*underdetermination*), survenire (*supervenience*), credibilitate (*reliability*); 5) Credințe, dorințe, speranțe, amintiri – atitudini mentale; 6) Imagine, reprezentare, inteligibilitate, 7) Simbol, concept, sens; 8) Personalitate, identitate, pluralitate; 9) Intensional, extensional; 10) Subconștient, emotiv, irațional; 11) Libertate, normativitate, subiectivitate.

Dintre *temele și problemele* frecvent abordate, enumerăm aici pe cele despre: natura eului, a conștiinței și a cogniției; realitate subiectivă, real psihic și mental; raportul minte-corp (mental-fizical, materie-conștiință, suflet-corp); definirea conștientului și a intenționalității; despre relația explicație-interpretare, logic-psihic, psihologie-știință empirică, empirism-hermeneutică; bazele biologice și psihologice ale cognitivului uman; gramatica mentală și structura limbajului; inteligență animală, umană și artificială; despre comunicare și experiment mental, libertatea voinței (liber arbitru) și determinarea deciziei; relația computer, self-inteligență, modele mentale; despre științele cognitive și cotitura psihologică în filosofie.

**Orientările principale prezente astăzi în aria filosofiei mentalului sunt:** *naturalismul, fizicalismul și funcționalismul, materialismul eliminativ și monismul anomalistic*, care susțin că mentalul face parte din lumea naturală, ca funcție a corporalului;



*mentalismul* și *holismul emergentist* după care existența mentală diferă de cea fizică și are o structură psiholingvistică și un comportament intențional ireductibil. Alte concepții frecvent dezbătute sunt: *conecționismul* – actele mentale sunt rezultatul conexiunilor neuronice interne cu structuri nodale paralele; *contextualismul* – actele mentale sunt rezultate ale influenței mediului asupra subiectului; *cognitivismul* – toate actele mentale sunt de natură cognitivă; *constructivismul* – factorii cognitivi și sociali interacționează în construcția actelor mentale. La acestea se adaugă *externalismul*, *reliabilismul* și *internalismul*, *neutralismul*, *teoriile identității*, *unității* și *constituirii* mentalului și fizicalului.

Ca **personalități** ce reprezintă în mod explicit această disciplină filosofică, îi putem enumera pe: Block, Churchland, Crane, Davidson, Dennett, Fodor, Hacker, Harré, Honderich, Mc Intyre, Mellor, Nozick, Papineau, Parfit, Pollock, Rorty, Ryle, Searle, Sellars, Sperry, Strawson și alții.

Apar astăzi și o serie de reviste de filosofie a mentalului și științe cognitive, dintre care amintim: *Mind*, *Mind and Machines*, *Philosophical Psychology*, *Philosophy and Psychology* etc.

După definiția dată de *Dicționarul de filosofie*<sup>2</sup> editat la Oxford, în 1996, **filosofia mentalului** este disciplina filosofică prin care se încearcă să se răspundă la întrebări cum ar fi: Este mentalul distinct de materie? Îl putem noi defini ca fiind conștient, și putem descoperi rațiunile principale pentru care decidem că și alte creaturi ar fi conștiente sau că mașinile pot fi construite în așa fel încât să fie conștiente? Ce înseamnă a gândi, a simți, a experimenta, a-ți aminti? Este de folos să separi funcțiile minții și memoria de inteligență sau raționalitatea de sentiment? Formează oare funcțiile mentale o unitate? Poate fi definit specificul evenimentelor mentale cu ajutorul conceptelor de intenționalitate, conștiență sau pe baza raportului minte-corp, formal-experi-mental, fizic-psihic?

În aria de interes a filosofiei mentalului se delimitează astăzi și disciplina **psihologiei filosofice** care se ocupă de fundamentele psihologice ale gândirii logice, de abordările metodologice ale psihologiei aplicate și cognitive, de natura psihicului, de aspectele etice ale psihologicului și relația lui cu neurofiziologicul, lingvisticul și sociobiologicul uman. Fiind și filosofie a științei, adică a

psihologiei, psihologia filosofică se referă și la corespondențele și diferențele între cunoașterea explicativă a științei fizice și cea hermeneutică a psihologiei mentalului, între obiectele și metodele științelor naturii și ale celor psihologice. Sfera psihologiei filosofice se întretaie cu aceea a filosofiei mentalului.

În afara abordării filosofice prezente în cele două discipline *mind* filosofia și *psihologia filosofică*, domeniul mentalului mai este cercetat, dar științific, de așa-numitele **științe cognitive** (de fapt, ale cognitivului). Aceste noi discipline s-au conturat interdisciplinar, reprezentând sinteze ale unor rezultate din cercetările neurologice, cibernetice, de teoria informației, psihologie, lingvistică, sociologie, antropologie.

**Științele cognitive** sunt definite ca fiind studiul sistematic al cognitivului uman și al fenomenelor asociate urmărind acumularea cunoașterii cu privire la comportamentul intențional mental pe baza faptelor rezultate din analogia cu mecanismele cibernetice. Se pornește de la observarea funcționării computerului și se constată că o serie de structuri și modele ale acestuia pot fi găsite și în cadrul proceselor din creier. Pe baza acestor *analogii, inverse* celor care stau la baza inteligenței artificiale (de la mentalul uman la programarea mașinii), au apărut noile științe cognitive și teoriile specifice acestora.

**Conecționismul** este una dintre teorii și consideră mintea ca fiind activitatea unei structuri neuronale invariante, un procesor neural, paralel distributiv. Conecționismul – reprezentat în principal de P. Churchland, S. Stich, P. Smolensky, reprezintă abordarea mentalului produsă de câmpul interdisciplinar al științelor cognitive în care intră elemente de psihologie, cibernetică, lingvistică, filosofie. În sistemul conecționist, spre deosebire de cel clasic, mintea reprezintă o structură neurală compusă din centre nodale multiple, legate între ele, fără a fi subordonate, însă, unui centru de procesare care ar conduce operațiile sistemului și fără a avea memorie precum au computerele convenționale. Conexiunile între centrele nodale sunt astfel realizate încât recrează informația de câte ori este nevoie. Conecționismul ar fi deci denumirea dată cercetărilor științifice din științele cognitive care intenționează să construiască astfel de modele de sisteme cibernetice, dar și

concepției filosofice care definește cunoașterea în astfel de termeni. Pornind de la studiul sistemelor de inteligență artificială și trăgând concluzii despre stările cognitive ale creierului uman, concepția conexiionistă favorizează teoria relațiilor psihofizicale și materialismul eliminativ. Se susține faptul că, dacă un astfel de model nou poate explica toate aspectele mentalului, atunci grupul de teorii clasice ale cunoașterii și entitățile postulate de ele ar trebui eliminate, întrucât conțin elemente obscure.

Concepția conexiionistă tinde să înlocuiască paradigma epistemologică clasică încercând să rezolve, pe baza unei noi structuri neurale, probleme ce apăreau ca insolubile și explicând funcționarea creierului asemenea unui transductor echipat cu posibilitatea reprezentării distributive și a detectării erorii. Pe această bază Paul Churchland<sup>3</sup> promovează așa-numita **teorie a materialismului eliminativ**. El consideră că psihologia simțului comun ar susține numai teorii false despre lucruri inexistente ontologic, cum ar fi: credința, dorința, intenția, amintirea, speranța, temerea, orice referire la ele rămânând o referire la ceva misterios și obscur. De aceea știința despre stările mentale se cere redusă, în viziunea sa, la abordarea bazelor neurofiziologice fizicale ale minții. În general, teoriile eliminativiste cer renunțarea la acele părți ale cunoașterii impregnate de confuzie și erori: în cazul filosofiei minții nu numai la psihologia bunului simț, ci și la termeni cum ar fi: „conștiință“, „eu“, „qualia“ (calități), „intenționalitate“ etc. Până în momentul în care științele cognitive se vor dezvolta în suficientă măsură încât vor reuși înțelegerea științifică a mentalului, arată Churchland, nici o descriere a „eului“ nu poate fi testată și, ca atare, astfel de încercări trebuie eliminate.

Întrucât științele cognitive au ridicat astfel de probleme filosofice cu privire la natura cunoașterii umane, filosofia minții și psihologia filosofică s-au considerat adesea exclusiv ca baze teoretice pentru dezvoltarea acestor noi discipline, ocupându-se de aspectele lor ontologice și epistemologice, cu natura și structura actelor mentale, cu relațiile lor cu contextul fizical (cum ar fi corpul uman și mediul natural), cu modelarea proceselor cognitive și reprezentative, cu definirea mașinii inteligente. Aceste tentative de a dezvolta o filosofie a minții inspirată de științele cognitive pe o direcție reducționistă sunt criticate ca

unilaterale de autori cum ar fi P. M. S. Hacker și T. Honderich. Ei arată că, în aria științelor cognitive, nu există evidențe destul de tari pentru a se vorbi de legi și generalizări de tip științific clasic, astfel că rămâne problema dacă științele cognitive pot fi un alt tip de știință, diferit de cele fizicale, și ontologic legitimat de un alt tip de existență decât cea real-substanțială. P.M.S. Hacker<sup>4</sup> – cunoscut comentator al lui Wittgenstein și promotor al unei concepții filosofice proprii asupra mentalului, afirmă că reducerea „mind filosofiei” la o simplă metodologie a științelor cognitive, constituie o nouă formă de barbarism „filosofic” (după cum a caracterizat-o și N. Malcom, discipolul lui Wittgenstein), inspirată în mare parte de scientism, dar din păcate de ceea ce are scientismul mai rău. În viziunea lui Hacker, filosofia mentalului și știința cognitivă au unghiuri diferite de abordare – filosofic și științific – a aceluiași obiect – mentalul.

Științele cognitive se ocupă de descrierea structurilor și proceselor mentale ca fiind distincte și de factorii sociali și biologici. Termenul de „știință cognitivă” a fost prima oară folosit cam la o decadă după cel de al doilea război mondial pentru a numi un mod de investigație psihologică în școala behavioristă, care dorea să evite introspecția subiectivă, având un interes explicit pentru cercetarea empirică a structurilor și proceselor mentale ce determină comportamentul agentului uman. Într-o astfel de accepție behaviorismul a fost reprezentat, în principal, de B.F. Skinner (1904–1990).

Astăzi științele cognitive sunt direct legate de existența computerului, un instrument cu utilizare dublă: pentru cercetarea realității fizice, dar și ca model al minții umane. Cu ajutorul lui se propune o nouă perspectivă explicativă asupra cunoașterii, care relevă dependența actelor cognitive de capacitatea informativă a subiectului asistat de calculator (amplificator și stimulator al inteligenței naturale). Se abordează astfel de teme clasice cu instrumentul oferit de Inteligența Artificială, favorizându-se apariția unui succesor modern pentru behaviorism, **funcționalismul**.

Principalele probleme care se pun la ora actuală în filosofia minții cu privire la științele cognitive<sup>5</sup> vizează orientarea funcționalistă și pot fi sintetizate în diverse întrebări cum ar fi:

1. Explicația formării credințelor: Ce este credința? Există ea real și cum poate fi explicată sau modelată ca atitudine propozițională? Se ocupă Fodor, Stich, Garfield.

2. Problema inferării stărilor interne pe baza unor comportamente externe observate: se poate ajunge la comportamente diferite pe baza acelorași circumstanțe interne, mai ales când e vorba de comportamente verbale, inferențiale sau introspective, deci intenționale? Au încercat răspunsuri Pylyshyn, Cummins și Dennett.

3. Bazele biologice ale mentalului și nașterea conceptelor: Care este natura proceselor cognitive, a percepției și reprezentării? Există abordări la Fodor, Putnam, Peacocke.

5. Conștiința și reflectarea: Ce înseamnă să fii conștient? Ce înseamnă reflectarea propriilor operații mentale și reprezentarea lor în concepte și credințe cu existență intențională? Au scris despre această temă Mc Intyre, Ornstein și Dennett.

6. Înțelegerea structurii gândirii prin structura limbajului natural, descoperită pe baza analizei lingvistice: Există o gramatică mentală? Are gândirea o structură sintactică ca și limbajul sau una semantică? Încearcă răspunsuri Fodor, Brand, Garfield.

## II. ORIENTĂRI ȘI TEME PRINCIPALE

Pe lângă fundamentarea metodologică a științelor cognitive, o altă preocupare principală a filosofiei contemporane a mentalului, care a provocat și ea dezbateri de anvergură este **problema reprezentărilor mentale**, cu temele asociate: natura conștiinței și a explicației psihologice, cauzalitatea mentală, mintea în calitate de computer (I.A.), raportul fizical-mental, intenționalitatea și experimentul mental.

Pe astfel de teme perspectivele reduționiste și non-reduționiste s-au succedat până astăzi, când o disciplină filosofică specială – filosofia mentalului (minții) și-a stabilit obiectul principal în cercetarea naturii conștiinței, a reprezentării și a tuturor stărilor mentale. În cadrul ei continuă însă disputa între fizicaliști și non-fizicaliști, între viziunile mecanic-deterministe și organic-finaliste.

După opinia lui Tim Crane<sup>6</sup>, o explicație a mentalului nu este altceva decât recunoașterea, în ultimă instanță, a faptului că mintea face parte din lumea naturală a cauzelor și efectelor, care este o lume ordonată de legi. Aceste legi sunt studiate de psihologie, așa cum legile fizice sunt studiate de fizică. Există deci o știință a naturii care se ocupă de mental, dar care nu este fizica, ci psihologia cognitivă. Ea are de înfruntat, dacă adoptă o viziune deterministă asupra spiritului, două probleme puse de acele două fenomene ce împiedică constituirea unei teorii explicative unitare asupra lumii ca întreg. Este vorba de **conștiință** și de **gândirea reprezentativă** care aduc în atenție două din temele esențiale ale filosofiei minții: cum poate un mecanism (chiar psihologic, nu fizic) să devină conștient de sine și a doua, cum poate el să gândească asupra, și să-și reprezinte lucruri din afară. Tema reprezentării stă deci în centrul atenției filosofilor mentalului, fiind legată direct de orice manifestare a spiritului uman. A spune și a scrie cuvinte, imaginile, semnele, simbolurile, gesturile, expresiile fetei sunt reprezentări, ele semnifică lucruri și evenimente din viața cotidiană. Problema este cum se reușește acest lucru? Pe de o parte, reprezentările sunt determinate de natură, pe de altă parte sunt ele însele structuri fizicale – vibrații ale aerului, mișcări, semne materiale? Reprezentările par naturale, dar din punct de vedere filosofic poartă o încărcătură misterioasă care combină conceptele de timp, adevăr și existență. Pentru a înțelege ce este reprezentarea trebuie cercetată starea reprezentatională a minții, legată direct de posibilitatea de reflectare și locul conștiinței în natură, de relația gândire-limbaj, de matura intenționalității și înțelegerii altuia, de posibilitatea existenței mașinii de gândit. Astfel, ajungem la cunoscuta **problemă mind-body** care se referă la modalitatea în care apare interrelația între lume și spirit, între minte și creier, între gândire și timp. Prezentă sub o formă sau alta în întreaga istorie a filosofiei cunoașterii, problema minte-trup a fost considerată de majoritatea gânditorilor ca insolvabilă. Odată cu nașterea științelor cognitive, a teoriei informației și a Inteligenței Artificiale tema s-a reluat sub toate aspectele ei metodologice și filosofice: dualiste, fiziciste, funcționaliste, epifenomenaliste.

Abordarea contemporană a raportului minte-trup nu a reușit să rezolve, totuși, probleme de tipul: cum poate mintea să își reprezinte ceva anume din afara ei și cum putem noi să aflăm

ceva despre procesul reprezentării? Încercăm să aflăm ceva despre mintea noastră prin introspecție, într-un fel diferit de cel în care încercăm să aflăm ceva despre cum gândesc și sunt ații, respectiv, prin observarea comportamentului lor (behaviorism).

Cu privire la raportul corp-minte se susțin astăzi o serie de concepții cum ar fi: 1) cele ale lui Feigl, Wiggins și Pollock, care acceptă că entitățile mentale intră în componența creierului, dar nu sunt identice cu el; 2) așa numitul „materialism eliminativ“ al lui Quine, Rorty și Churchland, care consideră că entitățile mentale trebuie eliminate din discursul rațional asupra creierului ca fiind neclare; 3) funcționalismul materialist reprezentat de Fodor, Putnam, Block, și Davidson care identifică stările mentale cu stări funcționale; 4) fizicalismul lui Smart, Wilson, Nagel care reduc mentalul la baza sa neurofiziologică; 5) concepția emergetist-dualistă-interacționistă, susținută de Popper și Eccles, după care eul conștient (lumea a treia) produce efecte verificabile empiric asupra lumii I-a (fizicalul) și asupra lumii a II-a (a stărilor psihice).

Problema *mind-body* mai este legată și de temele: imortalității (supraviețuirea sufletului după moartea fizică, dezbătută de Platon, Aquinas, Descartes, Ryle, Kant); **inconștientului** (văzut ca parte a activității mentale de Freud și alți psihanaliști) și **inteligenței artificiale** (relația între program ca soft-ul și hard-ul mașinii cibernetice). Ea apare ca una din problemele fundamentale ale filosofiei în general deoarece mentalul abordat ca substanță gânditoare, suflet, eu, natură inteligentă, psihic, spirit a fost întotdeauna pus în raport cu materia și lumea externă, cu corporalitatea, fizicul. De aceea în istoria filosofiei există numeroase orientări referitoare la relația minte-trup, minte-creier, dintre cele mai recente, o putem numi pe aceea fundamentată pe spiritul cuantic, care acceptă complementaritatea entității spirituale cu cea materială, trecând dincolo de monism și dualism, la o perspectivă holistă. Discuțiile despre natura activității mentale impun și întrebări cum ar fi: E posibilă comunicarea între două creiere deosebite? Ce au ele în comun? Cum poate un outsider înțelege alt creier? Entitățile mentale prezintă autoreferință și autoreproducere, de aceea se poate face analogia între programe de computer și coduri genetice mentale. Teoria Inteligenței Artificiale fundamentată pe ideile lui Church, Turing și Tarski demonstrează că pe baza programelor, mașina poate juca jocuri, poate demonstra teoreme și rezolva probleme, poate

compune muzică și folosi limbaje naturale. Mecanismele de declanșare pot produce reacții diferențiate, dar previzibile. Minte umană este tot astfel construită: la o anumită declanșare, un creier răspunde cam în același fel cu altul, de aceea copilul poate învăța orice limbaj. Inteligența poate fi considerată ca un exemplu de fenomen general al naturii. Mesajele ei, deși diferențiate, pot fi comunicate. Psihologia cognitivă se ocupă, la nivel științific, de modelul creier-minte-computer și de limbajul creierului, numit *mentaleză*. Mentea nu are dimensiuni și coordonate spațiale și nu este substanțială, dar se dovedește reală și naturală. Pe astfel de constatări referitoare la entitatea mentală și la cognitiv se bazează **orientarea naturalistă în filosofia minții**, care susține că entitățile fizicale produc cauzal stări cognitive tot atât de naturale ca și acestea. Se consideră că o astfel de concepție a produs o revoluție copernicană, întrucât a pus omul și cunoașterea sa în centrul Universului, obligând supranaturalul să se răstrângă tot mai mult, chiar din lumea spiritului. Naturalismul nu reduce lumea fizical-naturală la entități materiale, scrie Papineau<sup>7</sup>, ci o extinde la stările mentale, informaționale și conștiente chiar, la tot ceea ce științele pot explica, de la domeniile fizic, chimic și biologic la cele psihologic, sociologic, economic. Iar între ele, ca legătură, se poate vorbi de științe al căror obiect îmbină aspecte ale celor două domenii fizic și mental cum ar fi: neurofiziologia, Inteligența Artificială, științele cognitive. *Epistemologia naturalizată* s-a dezvoltat referitor la acest tip de cunoaștere a fenomenelor mentale care au aceeași calitate existențială ca și fizicul, dar pot manifesta și trăsături specifice ca survenirea (*supervenience*) și intenționalitatea.

Prin **survenire**<sup>8</sup>, se înțelege în filosofia mentalului o modalitate tipică de relație între anumite proprietăți care se nasc datorită altora. Stările psihice sunt stări relaționale, care survin din proprietăți ale fizicalului, ale creierului mai precis. Ele apar doar pe baza constatării că diferențele mentale nu pot fi detectate în cazul identităților fizicale. Doar diferențele în starea creierului nasc diferențe ale stării minții. **Intenționalitatea**<sup>9</sup> denumește o trăsătură exclusivă a fenomenului mental, aceea de a include cu intenție în sine un obiect, de a fi „direcționat” sau „a fi despre ceva”. Ea caracterizează stările conștiente, credințele, gândurile, dorințele, visurile, voința, toate fiind „despre” anumite lucruri.



**Analiza experimentului mental** constituie o altă temă importantă a filosofiei mentalului și a psihologiei filosofice. Cu privire la experiment există orientări care adoptă ideile *paralelismului*: Dumnezeu ar fi creat structura lumii analog cu cea a minții, astfel că imaginația înăscută a omului îi dă posibilitatea să preformeze în gândire scenarii ipotetice despre ceea ce s-ar întâmpla dacă o încercare ar avea loc în realitate. Cele mai puternice curente implicate în filosofia mentalului cu privire la experiment rămân însă cele *naturaliste*. Minteaa fiind o parte naturală a corpului uman se dovedește astfel construită încât poate să surprindă structurile lumii din care face parte. Cunoașterea reprezintă copierea și adaptarea „realului la real”. Se petrece același lucru ca și cu forma peștilor care parcă exprimă cunoașterea legilor hidrodinamicii, variațiile de formă confirmând capacitatea de selecție informativă. Explicația naturalistă arată că prin experimente se dobândește cunoașterea modală. Posibilitatea experimentului mental în studierea persoanei umane adevărește perspectiva naturalistă, dovedind că lumea mental-cognitivă și cea socială sunt tot atât de naturale ca și alte părți ale Universului și se conduc după legi ce pot fi descoperite de experimentele mentale – de la cele obișnuite, la cele cruciale. Există recompense biologice pentru adevăruri și sancțiuni (culminând cu distrucția) pentru erori, astfel că evoluția omului pare a se produce pe baza unor strategii cognitive bazate pe memorie, analogie, credințe, norme, opțiuni (dacă acționez într-un anumit mod, am anumite rezultate). Experimentul din lumea umană diferă de cel din lumea fizică doar prin faptul că cel din urmă nu poate provoca schimbarea structurilor de ansamblu ale lumii materiale. Experimentul social poate încerca stimularea și modificarea unui anumit tip de comportament prin propunerea unor noi modele de relații psihice sau cognitive.

Despre experiment și diversele concepții constructiviste sau naturaliste care apar în legătură cu el a scris Andy Pickering, care elaborează o teorie realist pragmatică a experimentului, îmbinând procedurile materiale cu modele conceptuale. Concepții care văd practica științifică drept o construcție psihică „manufactură a cunoașterii, act social contextual” au fost elaborate de Barnes, Bloor, Hacking, de Latour și Woolgar, de Lynch, Knorr-Cetina, Radder, Harré, Milikan.

Naturalismul și psihologismul secolului XIX care încercaseră să demonstreze ideea că există o bază structurală psihică pentru gândirea logică au fost resuscitate odată cu redescoperirea de către Wittgenstein a faptului că gândirea logică și chiar matematică se nasc și se dezvoltă în evoluția socială și psihică a ființei umane. Din acest moment se poate vorbi de o **cotitură psihologică în filosofie**, psihicul devenind obiect al disciplinelor filosofice. Cercetările asupra experimentului mental (numit, uneori, și psihologic) au contribuit în mod esențial la nașterea filosofiei mentalului și a psihologiei filosofice, întrucât ele au atras atenția asupra faptului că logica nu poate funcționa separată de psihologie, că orice act cognitiv cuprinde un obiect al cunoașterii și un subiect al său, dar și un al treilea factor contextual – inevitabil psiho-social, cu fundamente bio-sociale. Chiar în efectuarea experimentului sunt cuprinse astfel de aspecte contextuale ce apar în reproducerea artificială sau în modificarea intențională a unui fenomen natural, în scopul observării lui în condiții speciale, create deliberat de cercetător. O serie de idei ale adepților **epistemologiei naturalizate** se bazează pe observarea vieții reale din laborator, observație care demonstrează că faptele științifice produse prin experimente, deși create mental, nu pot fi decât naturale, pentru că înseși fenomenele mentale sunt naturale. Întrucât se bazează pe un model de tip cibernetic, experimentul mental se implică în mod esențial și în evoluția științelor cognitive.

Problematika filosofiei mentalului, sumar prezentată aici, cuprinde doar câteva din temele principale ale domeniului, mai ales cele care se află surprinse într-un fel sau altul în studiile din lucrarea de față. Este vorba despre: reprezentarea și 'cauzarea' mentală, conștiință și intenționalitate, formarea credinței (*belief*) și specificul științelor cognitive, bazele biologice ale comportamentului mental și cognitiv, gramatici mentale și lingvistice, dar, mai ales, despre raportul minte-trup și experimentul mental (analizate pe larg în studiile din partea a doua a acestui volum) de care se leagă direct problemele cotiturii psihologice în epistemologie, ale folk-psihologiei, ale teoriei Inteligenței Artificiale și conexiunismului, ale definirii naturaliste și funcționaliste a mentalului și fizicalului.

### III. PRECURSORI ȘI REPREZENTANȚI

După ce am prezentat în linii mari câteva teme și orientări teoretice ale filosofiei mentalului să ne întoarcem la precursorii și fondatorii teoriilor contemporane ale acestei discipline. Ne vom referi doar la câțiva: Reid, James și Wittgenstein, dintre precursori și la Ryle, Sellars, Davidson și Searle dintre fondatori.

**Thomas Reid** (1710–1796) este unul din principalii precursori ai filosofiei mentalului, care a scris: *An Inquiry into Human Mind on the Principles of Common Sense* (1764), *Essays on the Intellectual Powers of Man* (1785). El s-a ocupat de credință (belief) și de concept, dând intenționalității și conștiinței un rol central în interpretarea acestora. Orice credință presupune o operație a minții ca parte a existenței noastre naturale. Anticipând concepțiile contemporane cauzale, Reid susține că o viziune asupra credinței presupune conștiința activității noastre mentale și, ca atare, mintea umană se manifestă în mod esențial ca „metaminte“. El distinge între operația de a crede actul mental și obiectul credinței, conținutul ei. Se pot emite adevăruri despre lucruri care nu există, susține Reid, adevăruri despre non-existența lucrurilor, cum ar fi cele despre universale care ființează doar prin particulari. Ontologia sa particularistă se bazează pe existența exclusivă a individualilor. A concepe un universal înseamnă a concepe o semnificație în termeni generali.

În viziunea lui Reid, *doctrina conștiinței* reprezintă principala doctrină a filosofiei minții. Facultatea conștiinței este considerată ca puterea înăscută a minții de a performa anumite operații: a concepe și a crede, de pildă. Stările mentale sunt produse ale conștiinței, cel mai important aspect al acesteia fiind intenționalitatea (deși el nu o numește încă așa). Conștiința – ca meta-facultate – ne ajută să interpretăm semnificațiile lumii.

Reid face parte din școala scoțiană a filosofiei simțului comun, care s-a opus ideilor empirist-subiectiviste sau sceptice ale lui Hume și Berkeley. El crede că senzațiile primelor calități (Qualia) ale obiectelor ne vorbesc precum cuvintele, semnificând minții „percepții originare“, din care se nasc percepțiile calităților secundare.

Un alt precursor remarcabil ni se pare a fi: **William James** (1842–1910), psiholog și filosof, fondator al *pragmatismului*. El consideră că ontologic, lumea mea e alcătuită dintr-o substanță care nu e nici materială, nici spirituală. Existența în afara conștiinței nu are coerență, conștiința fiind un flux continuu, nediferențiat, de senzații și impresii, produs de stimuli externi. Adevărul reprezintă utilul.

Negând raționalismul, William James admite o metafizică a experienței, unde experiența științifică are același rol cu cea religioasă. El este împotriva asociaționismului în psihologie, opunându-i o „structură psihică organică“, ca precursora a behaviorismului. Operele sale principale sunt: *The Principles of Psychology* (1891), *The Will to Believe* (1897), *Pragmatism* (1907), *Pluralist Univers* (1907), *The Meaning of Truth* (1909). El intră în filosofie din direcția fiziologiei și a psihologiei și consideră că trebuie acceptate toate credințele, manifestându-se ca un antiintelectualist de tip bergsonian. Influențat de Locke, dar și de Edwards și Emerson, James consideră că principiile metafizice se bazează și pe sentimente. Dacă Locke vorbea de existența rațiunii intuitive a adevărului în experiență, James adaugă la experiență, emotivitatea. El consideră că există diferență între rațional și nerațional, dar nici principiile fizicii nu pot fi pur raționale. El apără liberul arbitru arătând că nici apelul la fapte nu asigură determinismul. James introduce atitudinile propoziționale: a crede, a nu crede, a te îndoii. Organicismul său se manifestă în psihologie, când se referă la tipuri diferite de credințe: provocate de experiența senzorială, rezultate din relații conceptuale, dar apărute și altfel, emotiv.

Un alt precursor al filosofiei mentalului este **Ludwig Wittgenstein**<sup>10</sup>, (1889–1951), despre care P.M.S. Hacker scrie (în *Wittgenstein: Meaning and Mind* și, apoi, în *Wittgenstein: Mind and Will*) că s-a ocupat de concepte psihologice cum ar fi: senzație, percepție, minte și trup, intern-extern, gândire-imaginație, identitate personală, conștiință, memorie și recunoaștere, voință, intenție, intenționalitate combinând interesul pentru filosofia limbajului cu cel pentru filosofia minții.

Spre mijlocul secolului, Wittgenstein făcuse deja pași importanți în domeniul filosofiei mentalului. Se poate spune chiar că

cea mai mare contribuție din a doua perioadă a activității sale, Wittgenstein a adus-o la constituirea filosofiei mentalului. El explorează rolul introspecției, al senzațiilor, intențiilor și credințelor în viața societății și caută să „submineze” viziunea carteziană în care subiectul era doar spectator al lumii. El afirmă ideea că limbajul privat și „jocurile de limbaj” domină activitățile de a ordona, a sfătui, a pretinde, a măsura și participă la constituirea „formelor de viață” prezente în aria de preocupări a filosofiei limbajului, dar și a mentalului. Filosofia sa este concepută ca o terapie a „eului” bolnav de tentația de a înlătura varietatea vieții mentale. Dacă în *Tratat*, Wittgenstein plasa limbajul într-o relație statică și formală cu realul, în *Philosophical Investigations* (1964) și *On Certainty* (1969), publicate postum, se ocupă de variația activităților, a formelor de gândire și ale limbii pe care le plasează în contextul formelor de viață. El susține că analogiile și aforismele pot folosi la rezolvarea vechilor probleme ale cunoașterii în modalități diferite de orientarea tradițională convențional-formală și lineară, ce caută generalitatea adevărurilor.

Negând interioritatea proceselor mentale Wittgenstein nu acceptă nici părerea contrară a posibilității exteriorizării și materializării sferei mentale. El nu propune înlocuirea imaterialului mental eterat și a mecanismelor oculte cu mecanisme tangibile și materiale, considerând că substituirea spiritului cu funcțiile sistemului nervos central nu conduce la înlăturarea confuziilor cu privire la natura evenimentelor mentale. Wittgenstein îi reproșează lui Mach de a fi prezentat o imagine a câmpului vizual în care existența și poziția ochiului și faptul că ochiul produce câmpul vizual într-o anumită manieră sunt elemente determinante pentru caracteristicile imaginii: prezența unor obiecte distincte, proporțiile obiectelor, perspectiva. Presupunând că văd un obiect determinat în câmpul vizual, de pildă nasul meu, scrie Wittgenstein, este evident că altul nu vede acest obiect în același fel. Aceasta nu înseamnă că spațiul vizual îmi aparține, fiind deci subiectiv. Obiectele spațiale au prerogative egale, ele sunt însă surprinse diferit.

Fiind împotriva *behaviorismului* și a *materialismului eliminativ* Wittgenstein acceptă indeterminismul și incertitudinea cu privire la interioritatea umanului. Calitatea fenomenologică ime-

diată a senzației interne se dovedește incomunicabilă, fapt ce duce la scepticismul de toate tipurile. Atunci când omul a devenit conștient de sine a aflat că „eul“ este aici și acum. Apariția conștiinței legată de separarea obiect-subiect, prin spațializare și localizarea eului a creat posibilitatea solipsismului. Wittgenstein încearcă să-l infirme prin acceptarea „ochiului geometric“ independent de ochiul „eului conștient“. El arată că problemele oricărei orientări reducionista asupra gândirii și a eului uman (a behaviorismului în primul rând) se nasc din imposibilitatea de a face față holismului mentalului, adică, faptului că modul în care acționează și se comportă o persoană nu este în funcție de o anumită dorință sau credință ci, de un întreg câmp structurat de credințe și dorințe.

Gilbert Ryle (1900–1976) <sup>11</sup> stă fără dubiu pe primul plan între cei care au fundamentat filosofia minții din perspectivă behavioristă. Ryle a învățat mult și de la Wittgenstein, asimilând idei ale acestuia. Preocupările sale cu privire la limita între sens și nonsens și la transgresiune se originează în citirea *Tratatului* și au fost îmbogățite prin întâlnirea cu Wittgenstein, în 1930. Distanța parcursă de Ryle, cu sau fără asistența lui Wittgenstein, de la lucrarea sa *Systematically Misleading Expressions* (1923) prin *Philosophical Arguments* (1954) la *Dilemmas* (1954) s-a dovedit foarte mare. Concepția filosofică pe care a propus-o și a exemplificat-o în *The Concept of Mind* cu privire la existența unei „geografii logice“ a conceptelor, dovedește afinitatea cu ideea lui Wittgenstein despre o reprezentare dirijată a folosirii cuvintelor. Insistența lui Ryle în *Proofs in Philosophy* (1954) asupra concepției că filosofia conține afirmații și respingeri, dar nu teoreme compuse din premise de necontrazis, nu pare nici ea a fi una pe care Wittgenstein ar fi negat-o. Ryle nu obiectează la ideea de a vorbi despre „teorii“ în filosofie, ci susține că nu poți fi sigur că teoria filosofică este la fel cu cea științifică. Diferențele față de Wittgenstein sunt superficiale, dar ele indică imposibilitatea lui Ryle de a aproba în întregime concepția acestuia cu privire la propozițiile gramaticale și la autonomia gramaticii.

Încrederea lui Ryle în ideea că rădăcina erorii existente în concepția carteziană a minții a fost o categorie-greșală apare, după cum relevă și Austin, ca o suprasimplificare atât a surselor

mitului dualismului cartezian cât și a căilor de respingere a numeroaselor lui confuzii. Concepția lui Ryle asupra categoriei-eroare era prea slab construită ca să suporte greutatea pe care chiar el le-a pus asupra ei. Ryle crede că *behaviorismul* reprezintă cea mai bună apărare împotriva cartezianismului întrucât explică mentalul prin dispoziții de comportament. Astfel că Gilbert Ryle înlocuiește în lucrarea sa *The Concept of Mind* (1949) „ce cunoaștem“ (knowledge that) cu „cum cunoaștem“ (knowledge how). El încearcă să înlăture concepția carteziană a separației duale mind-body și consideră că ceea ce apare în interiorul corpului ca a fi mental este un fel de „strigoiul din mașină“, un fel de categorie greșită care prezintă faptele vieții mentale ca aparținând unui anumit tip de categorie, când, de fapt, ele țin de altele. El crede, ca și Wittgenstein, că eroarea apare datorită incapacității oamenilor de a folosi corect termenii, ca de pildă „spirit“ sau „mind“, pe care unii îi consideră analogi cu „corp“. Când vorbim de capacități mentale nu ne referim la procesele ascunse din creier, ci la acte externe observabile pe care le producem în mod inteligent. Ryle critică legenda intelectualistă care consideră că a ști *cum* înseamnă a ști *ceva*. Cunoașterea poate fi descrisă doar în termeni de capacități și dispoziții. A ști este un verb al capacității care arată că o persoană poate acționa corect. A crede este verb tendențial, cuvânt motivațional. Noi ne întrebăm *de ce* o persoană crede *ceva* și *cum* o persoană cunoaște *ceva*. Capacitățile au metode, pe când credințele, obișnuințele și înclinațiile au surse.

Wilfrid Sellars (1912–1989)<sup>12</sup> – filosof american – se consideră el însuși a fi părintele *funcționalismului mental* (*Empiricism and the Philosophy of Mind* – 1956). Explorând natura gândirii și a experienței el apără concepția gândirii ca limbaj intern și atacă teoria datului în toate formele sale. Toate evenimentele intenționale există pentru el în spațiul motivațiilor raționale, chiar o experiență pură reprezintă un concept teoretic, fiind reflectarea gândurilor noastre asupra lumii și asupra autorității cu care impunem imagini și concepții despre un „eu“ conceput ca fiind dotat natural cu intenții, semnificații, coloraturi specific umane.

W. Sellars este preocupat să păstreze o viziune care concepe omul ca agent rațional autonom, ireductibil ca personalitate. Imaginea științifică a omului trebuie să includă după părerea sa

aspectele libertății și ale intenționalității pe care kantianismul le exclusese. Sellars subliniază că numai funcțional, nu și exten-sional-ontologic, stări ale creierului pot fi echivalate cu intenții, dorințe, gânduri omenești. Sellars, deși acceptă fizicalismul, îl separă de cunoașterea personalității. El manifestă interes pentru funcționalism, așa cum apare acesta la H. Putnam, J. Fodor, M. Block. După părerea sa structurile fizice materiale determină funcția intențională a mentalului. Cedând reducionismului, el prezintă senzațiile ca nefiind altceva decât stări ale memoriei pe care funcționalismul le localizează. Conștiința apare la Sellars ca produs al unor structuri cu funcții semantice. Ca atare, imaginea științifică a omului se obține prin transferarea mentalului (experiența conștientă) într-o a treia lume, cea a exprimării prin limbaj. Sellars a încercat în concepția sa despre „republica existențelor raționale“ să rezolve problema ireductibilității agentului uman dotat cu calități deliberatoare.

**Donald Davidson** (1917)<sup>13</sup>, în programul său de filosofie a mentalului, caută elaborarea unei teorii unificate a semnificației și a acțiunii care nu este altceva decât o concepție filosofică preocupată de adevăr și interpretare, dar și de acțiuni și evenimente mentale. Lucrările sale principale sunt : *Essays on Action and Events* (1980) și *Inquires into Truth and Interdependence* (1983). El conectează viziunile asupra mentalului cu analizele comportamentului lingvistic și nonlingvistic explicând și cum anume am putea fundamenta filosofia minții fără a postula un obiect al conștiinței imediate. Davidson se ocupă de natura gândirii susținând o poziție externalistă după care cunoașterea altor minți nu trebuie să se bazeze pe cunoașterea empirică a mentalului altuia. Se poate face un pas în determinarea conținutului gândirii altcuiva prin descoperirea a ceea ce cauzează acele gânduri. Nu se poate identifica, susține el, gândirea individuală dacă nu există și alte creaturi gânditoare aplecate asupra aceleiași lumi naturale. Mintea și părțile ei constitutive au următoarele proprietăți: identifică sau ajută la identificarea gândirii prin conținutul ei, delimitează aspectele esențiale ale psihologiei și logicii ei care pot fi cunoscute de altă persoană care gândește. Davidson arată că, deși ar trebui să căutăm obiecte ale cercetării despre care erorile să nu fie posibile, nu este posibil să găsim așa ceva întrucât erorile au



substratul în subiect. Davidson se ocupă de raportul între gândire și limbaj și, în legătură cu aceasta temă gnoseologică, de modul în care persoana poate să se autoidentifice ca ființă gânditoare. Natura omului și esența sa se cer însă căutate și dincolo de studiul limbajului, în cercetarea interacțiunii între ființele care identifică și răspund similar la aceiași stimuli, adică într-o relație psihologică. Donald Davidson consideră că în ce privește explicarea genezei gândirii logice, teoriile externaliste și internaliste referitoare la cunoaștere trebuie să abordeze în mod complementar anumite cupluri ca acela al erorii și adevărului, sau al descrierii și opiniei etc.

*Monismul anomalistic*, cum își denumesc D. Davidson concepția, respinge ideea că toate fenomenele mentale ar fi explicabile pur fizic. Unele evenimente mentale și personalitatea ca o grupare a lor sunt, după părerea sa, „anomalii” față de explicațiile științei fizice. Nu poate exista ca atare o abordare total deterministă, cauzală a personalității. Legile științifice sunt „homonomice”, pe când cele psihologice sunt „heteronomice”. Vocabularul referitor la mental și la fizic diferă. „Holismul mentalului” nu poate fi surprins de teoriile fizice. Ar fi eficace să se poată elimina, cum cere Quine, din fizică, biologie și psihologie, orice intenționalitate, numai că, după cum demonstrează Davidson, această pretenție este de nesusținut când este vorba de personalitate, adică de om. *Intenționalitatea* face cunoașterea mentalului „heteronomică”. Limbajul științei exacte explică schimbarea fizică prin condiții inițiale și legi, limbajul psihologiei explică acțiunea umană prin „background”-uri, opinii și intenții. În ultimă instanță, Davidson dă o soluție kantiană. Persoana nu poate fi pe deplin analizată în termenii științei, dar nici cunoașterea nu are ca singur ideal „homonomia” predictivă și, ca atare, nu este reductibilă la cea științifică. Imaginea științifică a omului s-ar obține prin considerarea mentalului (experiența conștientă) ca o anomalie și transferarea lui într-o a treia lume, aceea a exprimării prin limbaj. Prin „idealul constitutiv al raționalității” Davidson încearcă să se opună reducționismului.

După John Searle (1922)<sup>14</sup> filosofia analitică a produs cu privire la mental orientări de tipul behaviorismului și al funcționalismului, ambele menținute în atenție și astăzi de speranțele produse de științele cognitive și de temerile născute de metafizica

carteziană. De la Ryle la Rorty filosofia analitică a adus prejudicii mentalului pentru că l-a caracterizat ca posesor al unor caracteristici misterioase și oculte, scrie Searle, susținând că a observat acest lucru când a încercat să extindă analiza actelor de vorbire la stările intenționale: credințe, temeri, speranțe, dorințe, experiențe vizuale. Au apărut imediat îndoieli „ontologice“ asupra încercării lui din partea analiticienilor. Etapele pozitivistice și verificacioniste ale filosofiei analitice păstrează, crede Searle, tarele behaviorismului chiar atunci când îl declară mort, întrucât caută verificarea propozițiilor despre mental în observarea comportamentului celorlalți. De același fel de defecte suferă și funcționalismul dezvoltat de fizicalism, care susține teoria identității stărilor mentale cu cele neurofiziologice. Materialismul funcționalist devine materialism eliminativ când exclude din obiectul său proprietățile ireductibile ale mentalului. Se încearcă definirea mentalului prin funcțiile sale sau prin relații cauzale. Funcționalismul a evoluat spre o perspectivă organică bazată pe modelul mașinii Turing în viziunea căruia stările mentale nu mai sunt stări funcționale de orice fel, ci stările logice ale computerului, procesele mentale fiind descrise ca procese computaționale de care se ocupă psihologia cognitivă și Inteligența Artificială. H. Putnam și D. Dennett susțin acest tip de funcționalism.

Searle arată că în orientările funcționaliste apar o serie de probleme ca de pildă confuzia între întrebarea: Care sunt proprietățile intrinseci ale stărilor mentale și întrebările: Când putem atribui astfel de stări mentale unui sistem? și În ce condiții suntem îndreptățiți să atribuim astfel de atribute intenționale? Este vorba de confuzia între a putea clarifica ce este o stare mentală și între a putea examina modul în care a treia persoană atribuie o astfel de stare mentală altora. Altă defecțiune a funcționalismului vine din concepția că științele cognitive și computerul ar permite să pătrundem în interiorul operațiunilor minții. Adică se crede că problemele filosofice ale mentalului pot fi transformate în probleme tehnice și științifice care vor fi rezolvate de programele cibernetice. Searle arată că toate aceste păreri sunt false din două motive: unul, programul cibernetice nu poate garanta prin sine forța cauzală specifică biochimiei cerebrale, al doilea, un sistem ca agentul uman poate instanția un program

fără să aibă stări mentale relevante. Searle crede că teama de a accepta în spatele aspectelor științifice, elementele metafizice ale naturii entităților mentale duce la o falsă punere și a problemei mind-body (minte-trup) și a celei a percepției altor minți.

Searle consideră că agenții umani pot produce evenimente, deci există cauzare mentală sau intențională. Diferența teoriei sale față de teoria lui Hume a cauzalității constă în faptul că dacă la Hume nu există experiență a cauzatorului, la Searle cel ce acționează sau percepe lumea are o stare auto-referențială astfel că relația cauzală apare ca parte a conținutului, nu ca obiectul experienței sale. Searle crede că lumea conține regularități cauzale care pot fi descoperite, dar acceptă că ele pot fi independente de experiențele noastre. Noi pornim de la prezumția că lumea are un grad de regularitate, prezumție care se înscrie în background-ul pe care Searle îl definește ca „un set de capacități reprezentanționale non-mentale care ne permite să ne reprezentăm participarea“. După el există o cauzalitate intențională care este însă o subclasă a cauzalității în general.

#### IV. INTERFERENȚE CU ALTE DISCIPLINE FILOSOFICE

Filosofia mentalului și psihologia filosofică, metodologia științelor cognitive de care ne ocupăm pun într-o lumină nouă modalitățile de cunoaștere a ființei umane și interferează cu alte discipline filosofice care se referă la esența și existența „eului și a stărilor sale cognitive“. Există două probleme în acest areal, diferite, dar înrudite și tot atât de actuale. Una se referă la mutația produsă în câmpul epistemologiei odată cu includerea în obiectul acestei discipline filosofice a preocupărilor pentru om, ca urmare a acceptării entităților istorice și antropomorfe în structura propriu-zisă a demersului cognitiv, acceptare obținută pe baza relevării dimensiunii istorico-psiho-sociale a genezei și evoluției științei, a naturii metaforice și analogice a limbajelor și modelelor de cunoaștere, a rolului intreselor și valorilor nu numai în procesul desco-

peririi, ci și în cel al argumentării și validării cunoașterii. Cotitura epistemologiei spre abordarea mentalului nu este străină de poziția actuală a unor discipline, precum psihologia, sociologia, antropologia, axiologia, care au demonstrat pregnanța determinărilor contextuale și valorice ale demersului cognitiv și au contribuit la definirea științei ca formă a existenței socio-culturale a omului în Univers. Cea de-a doua problemă vizează caracteristicile actuale ale domeniului filosofiei științelor despre om, în care s-ar include științele cognitive, adică filosofia referitoare la acele științe care, în clasificarea lui J. Habermas<sup>15</sup>, constituie categoria științelor hermeneutice, a treia categorie de științe alături de categoria științelor empirico-analitice și de cea a științelor praxiologice (medicina, economia, sociologia, politica). Științele hermeneutice se caracterizează prin faptul că urmăresc înțelegerea și consensul intersubiectiv, în vreme ce celelalte caută cunoaștere nomologică sau respectiv rezultate performante. Unele abordări susțin că numai o parte din științele sociale, și anume cele care cer un mod de interpretare neutilizat în științele naturii, ar fi științe umaniste. Este vorba de modul de interpretare autoreflexiv care li se impune acestora la trecerea așa-numitei bariere ontologice ce separă fizicul de mental, existentul de conștient, descriptivul de normativ, mod prin care științele despre om se apropie de specificul disciplinelor filosofice.

Astfel de precizări comparative cu privire la obiectul filosofiei științei și cel al filosofiei mentalului se dovedesc foarte importante în condițiile în care, se recunoaște astăzi că toate științele presupun un anumit tip de interpretare și se vorbește atât de mult despre „antropologizarea“ științelor naturii și „naturalizarea“ științelor sociale, despre un reduționism parțial și despre un reduționism „revers“ (adică folosirea de metode proprii științelor sociale în științele naturii (vezi și utilizarea rezultatelor ciberneticii în cercetarea minții omului). Nașterea unei mari varietăți de logici face mai plauzibilă discuția despre deosebiri graduale, și nu dihotomice, între științele psiho-sociale și cele ale naturii. Pe baza analizei statutului logic al metodelor de generalizare, de argumentare și de verificare au apărut de curând încercări de apropiere chiar între discipline îndepărtate, cum ar fi, de pildă, psihologia și filosofia fizicii. Conștientizarea faptului că personalitatea umană și produsele creației sale sunt ireductibile

la fizicalism, conduce la ideea statutului diferit al cunoașterii în disciplinele socio-umane între care se includ în științele psiho-cognitive față de cunoașterea naturii. După părerea unor gânditori, precum M. Foucault și I. Buchler sau M. Eliade și P. Ricoeur, aceste științe ar trebui să se ocupe de „integrități” și nu de „identități”, de hermeneutică și nu de descripție. Toate structurile create de om – se susține din această perspectivă – sunt complex integrate, semnificative, și se succed distinct. Există „unități de discurs” și „unități de explicație” în istoria intelectuală, prin care arta, știința și filosofia se înscriu în spațiul epistemic. Complexele de semnificații rezultate din activitățile artistice, științifice, filosofice, constituie „unități perspective”. Omul se dovedește ireductibil cel puțin bidimensional, ca existență naturală și psiho-culturală, iar problematica epistemologică a umanului presupune cu evidență problematica ontologică a mentalului.

**Paul Ricoeur**<sup>16</sup>, reprezintă unul din principalele curente filosofice contemporane: hermeneutica. Lucrările sale încearcă o reconstrucție teoretică a sensului termenului „Eu”, plecând de la analiza unor concepții fenomenologice, pragmatice, semantice, cu privire la om. Ricoeur arată că predicate fizice, psihice, lingvistice desprinse din spațiul natural și social caracterizează persoana ca existență autoreferențială și ca „Eu în acțiune”. Ilustrativă pentru evoluția concepției lui Ricoeur se dovedește propunerea sa cu privire la construirea unei paradigme a proiecției eului de către o disciplină filosofică „mixtă”, care să conțină elemente fenomenologice și analitice și să folosească modelul explicație-înțelegere-interpretare pentru a unifica dimensiunile ontologică, axiologică și lingvistică ale eului care acționează. P. Ricoeur plasează filosofia limbajului în cadrul filosofiei acțiunii. După părerea sa, asimetria între eul producător de acțiune și eul receptor al ei, impune introducerea dimensiunii axiologice în orice structură teleologică. Paradigma proiecției sinelui pune problema intenționalității, iar starea morală a ipseității nu poate să nu influențeze structura dialogică, raportarea la celălalt. Întrucât conștiința morală, subliniază pe bună dreptate Ricoeur, constituie interiorizarea celui alt, responsabilitatea față de alter-ego ea se implică în toate științele sociale.

O altă orientare epistemologică direct legată de filosofia mentalului pornește din cadrul Școlii de la Frankfurt. Dintre reprezentanții ei, **J. Habermas**<sup>17</sup> a fost consacrat în domeniul filosofiei științei, printr-o originală concepție asupra intereselor cunoașterii, opusă viziunii neopozitiviste. În unele din lucrările sale, el se ocupă de interpretarea individualității ca sistem comunicativ și acțional și de filosofia științelor sociale în contextul criticilor contemporane aduse modernismului de Foucault, Derrida sau Lyotard. După părerea sa, problemele apărute în legătură cu existența dimensiunilor simbolice și subiective ale vieții sociale, cu interrelația dintre acțiune și mijloacele de comunicare, cu înțelegerea rațională a istoriei și autoînțelegerea subiectivă a persoanei indică limitele hermeneuticii, ale funcționalismului, ale teoriei sistemelor și ale concepției lui Popper, atunci când acestea se referă la științele sociale. El atrage atenția asupra importanței interlocutorului, a persoanei a doua, în orice încercare de identificare a ființei umane ca subiect comunicativ și activ, a necesității stabilirii prin deducție universală a principiilor morale și a structurii semnificațiilor lumii trăite a celui alt. Habermas accentuează necesitatea preocupărilor pentru interese și valori în orice cercetare a psihicului și a socialului. În contrast cu absolutismul și pozitivismul, Habermas susține filosofia contextualist-construcțivă după care cunoașterea este întotdeauna cunoaștere produsă de subiect și are un statut dependent de „condițiile subiective” în care intră și de faptul că subiectul devine conștient de contextul constitutiv al cunoașterii sale. Locul și specificul disciplinelor socio-umane, între care se află și psihologia cognitivă în câmpul cunoașterii științifice, constituie astăzi unul din obiectele principale ale filosofiei științei, abordarea lor contribuind la conturarea caracteristicilor și metamorfozelor din filosofia contemporană a științei. Nașterea unor noi tipuri de epistemologie – naturalizată, evoluționistă, constructivistă – se dovedește a fi fost determinată de o anumită cotitură psiho-sociologică, ce evidențiază tendințe reificatoare în autoconștientizarea subiectului cunoscător, determinarea biologică, psihologică și sociologică a abordării obiectului natural în cunoaștere, conceperea obiectului științelor despre om ca fiind o realitate cu multiple structuri posibile. Problema specificului științific natural sau

umanist al psihologiei cognitive se leagă direct de amintita cotitură. Principalele orientări filosofice ce se înfruntă pe acest teren sunt astăzi hermeneutica, personalismul, pragmatismul, naturalismul, identificarea persoanei umane ca obiect și subiect al cunoașterii fiind pentru ele, după câte se pare, una dintre cele mai dificile teme. *Teoria substanțialistă a identității* persoanei propune egalizarea, simetria între „actor“ și „receptor“, între „cel ce face“ și „cel ce receptează“ acțiunea. *Teoria relativistă* susține asimetria între cele două părți și posibilitatea „victimizării“ receptorului în cazul în care dimensiunea axiologică este pierdută din vedere. În structura cunoașterii persoanei se relevă trei tipuri de experiență: a eului propriu (autointuirea), a celuilalt (intersubiectivitatea) și a lucrurilor (transubiectivitatea), în cadrul concepțiilor substanțialismului clasic, relativismului modern și a structuralismului contemporan. Limitele acestor concepții asupra persoanei au condus la nașterea unui personalism integral moral, care consideră că orice cercetare științifică a ființei umane trebuie să implice dimensiuni cognitive, biomedicale, psihice, juridice și morale. Această arie tematică cuprinde și științele cognitive despre care se crede că ar trebui să se ocupe de gândirea omului în mod analog cu computerul, adică decupând neutral actul cognitiv din contextul său biologic și social, inteligența artificială și stările mentale umane ar trebui modelate cu aceleași metode excluzând intenționalitatea. În acest punct apar din nou marile discuții cu privire la diferența mental-fizical și la necesitatea luării ei în considerație, în cadrul cunoașterii socio-umane. În legătură cu problema alienării minții față de corp, a gândirii față de realitate, pe care au relevat-o printre alții Marx și Whitehead, s-au născut și încercările actuale de definire a eului prin „decentrare tematică“. Apărută ca trăsătură a postmodernismului, multiplicarea perspectivelor accentuează interacțiunea între persoane și lucruri, și conduce la ideea că realitatea poate fi cunoscută doar în pluralitatea interpretărilor sale. Presupunând un nou raport între realitate și ficțiune, descriere și interpretare, cunoaștere și contexte, noua perspectivă influențează orientările din filosofia științelor despre om, din filosofia limbajului, dar și din filosofia mentalului.

Un exemplu interesant ni-l oferă poziția lui **R. Rorty**<sup>18</sup> care, venind dinspre școala analitică, aderă la postpozitivism. Concepând

limbajul drept o parte a comportamentului ființei umane, Rorty respinge realismul, esențialismul și reprezentationalismul deoarece ele susțin doar existența funcției reflectorii a limbajului. Propunând o soluție pragmatică opusă realismului, Rorty împrumută definiția relaționistă a lui Sellars pentru filosofie: o încercare de a înțelege modul în care lucrurile (în cel mai larg sens al termenului) se leagă împreună (în cel mai larg sens al expresiei). Filosofia, în viziunea sa, se bazează pe credințe și tipuri de limbaj.

Cunoașterea nu este redarea adevărată a realității, ci obținerea de obiceiuri și acțiuni care fac față realității, scrie Rorty. Există diferențe sociologice, dar nu epistemologice între matrici disciplinare cum ar fi fizica teoretică și critica literară. Principalii autori pe ale căror idei își fondează Rorty orientarea sunt Wittgenstein, Heidegger și Dewey. În *Filosofia și oglinda naturii* m-am ocupat mai ales de Sellars și Quine, iar în ultimii ani am citit mai ales din Davidson, Putnam și Papineau, notează el, încercând să demonstreze că „teoria cunoașterii spectator” se cere abandonată de o societate democratică. Astăzi *limba* a înlocuit *mintea* în raportarea la realitate. Deci, discuția a trecut de la problema dacă lumea reală este independentă de minte, la problema care propoziții adevărate (dacă există vreuna) stau în relație reprezentatională cu sistemul nonlingvistic. Ca antirepresentationalist Rorty susține că limbajul nostru, ca și corpurile noastre, sunt produse de mediul în care trăim. Mintea și limbajul nostru nu sunt mai mult în afară de realitate decât corpurile noastre. El neagă faptul că doar unele iteme ale limbajului sau ale minții „corespund” sau „reprezintă” realitatea (mediul), pe când altele nu.

Interpretările lumii reale nu pot fi decât multiple și divergente, în artă ca și în știință, susțin textualiștii postmoderniști<sup>19</sup>. Critici ai perspectivei postmoderniste arată însă că, după ce analiza textuală aplicată la cercetarea acțiunii umane a demonstrat existența a numeroase modele metaforice interpretative, sunt necesare: o bază logică de explicare a acestei multiplicități, o structură logică obiectivă, în raport cu care s-ar fi diferențiat aceste modele ale actelor și „actorilor” sociali. Întrebarea ce rămâne se pune în legătură cu poziția de pe care se judecă diferențele: este aceasta poziția unei ființe raționale universale sau a unei ființe aparținând la o anume cultură și societate?



Filosofia mentalului interferează evident cu disciplinele care se ocupă de modelele umanului stabilite la nivelul limbajului, dar și cu cele fundamentate la nivel socio-psihologic și chiar antropofiziologic. Geneza raționalității umane s-a produs prin conturarea unor trăsături specifice: limbajul, conștiința, moralitatea în cadrul procesului evoluției. În ultima vreme se poartă numeroase discuții cu privire la modalitatea în care aceste trăsături sunt investigate de disciplinele biologice, psihologice și de cele socio-umane. O serie de descoperiri recente și puternicul avânt luat de cercetările biologiei genetice au adus în atenție și au descris, în termenii evoluționismului, fenomene umane și patternuri mentale aparținând culturalului, socialului și psihologicului. În anii '70 au apărut lucrări (Th. Nagel, G. Herman) care puneau problema fundamentelor personalității culturale, făcând analiza perspectivei lui Hume sau Kant cu privire la esența umană și referindu-se la comportamentul natural al individului. O serie de alți gânditori s-au ocupat, odată cu dezvoltarea etologiei, de comportamentul uman prin analogie cu cel animal. După unele opinii, la acest nivel, comparativismul s-ar fi manifestat doar în ceea ce privește forma comportamentelor. Începând din 1981, însă, au apărut teorii care susțin nu numai posibilitatea comparațiilor formale între lumea animală și lumea umană, ci și determinarea conținutului comportamentului uman prin elemente ce fac parte din structura biologicului în general. O astfel de teorie este cea a coevoluției genetice și culturale a lui Ch. J. Lumsden și E.O. Wilson<sup>20</sup> sau cea a lui R. Dawkins<sup>21</sup> cu privire la predeterminarea genetică a existenței culturale. Fără a se delimita de antropologia fiziologică, ce răspunde la întrebarea ce poate face natura din om, de antropologia filosofică ce răspunde la întrebarea, ce poate și ce trebuie să facă omul din el însuși, această din urmă teorie subordonează evoluția culturală, evoluției genetice. Astfel, pentru a explica evoluția mentală, Dawkins face apel la teoria variațiilor rapide a lui Lamarck; transmisia culturală s-ar dovedi similară cu cea genetică, având loc prin selecția unor unități de informație prezente în mintea omului – „menele” – analoage cu genele, venind de la nivelul subconștientului colectiv. Ideea se apropie de cea a psihologiei arhetipale, care susține prezența unui substrat psihic ancestral suprapersonal în fiecare om.

După părerea altui biolog, R.D. Alexander<sup>22</sup> nu există comportamente umane pur culturale, după cum nu există nici unele pur genetice, mediul influențează expresia oricărui comportament, iar cultura poate fi decisivă în evoluția biologică, asemenea unei părți a mediului. Așa cum fizicienii caută un principiu unificator al universului, cercetătorii universului uman ar trebui să caute și ei baze unificatoare. Dacă fizicienii nu stărnesc râsul în demersul lor unificator, nu există nimic ridicol nici în cercetările biologilor de a explica umanul. Înțelegerea unor fenomene biologice ca îmbătrânirea, selecția sexuală, grija părintească, relațiile de rudenie se dovedește necesară pentru înțelegerea naturii rațiunii și conștiinței umane, a conflictelor de interese, a rolului moralității și eticii. Orice persoană, susține Alexander, este programată de istoria selecției naturale pentru a-și asigura supraviețuirea materialului genetic prin reproducere. Pentru aceasta omul are nevoie de resurse, de beneficii. În lupta pentru existență s-au născut de-a lungul timpului grupurile umane competitive și agresive unul față de altul. Nevoia de apărare împotriva altor grupuri a dus la nașterea, în evoluția speciei, a trăsăturilor prevederii, conștiinței, inteligenței, care ar explica natura sistemelor morale și comportamentul indivizilor în cadrul lor. În lucrările sale, dincolo de numeroase investigații amănunțite ale relației dintre biologic și moral, R.D. Alexander a reușit să demonstreze că nu este necesară aura de mister în abordarea conceptului moralității și în interpretarea sistemelor morale, problemele morale și etice fiind direct legate de conflictele de interese din societate.

Cu privire la socio-biologie, Alexander susține, pe bună dreptate, că aceasta nu este o disciplină separată, întrucât se constituie ca o parte a biologiei. În momentul când, ca și etologia (întemeiată de K. Lorenz), socio-biologia s-a transformat în filosofie, la interpretarea problematicii ei s-au angajat nespecialiști, care folosind argumente neștiințifice au provocat o mare dispută între adepți și oponenți. Pentru biologi, termenul socio-biologie nu este controversat, pentru că el denumeste preocupările pentru genetica și fiziologia comportamentului uman. Mulți autori consideră biologicul ca ceva opus psihologicului, socialului și culturalului; de fapt biologii studiază în mod curent, după Alexander, ceea ce umaniștii numesc psihologic și cultural, atunci când ei se ocupă de cauzele și stimulii psihologici, sociali, culturali ai comportamentului.

Biologia, în concepția lui Alexander, reprezintă știința viului și tema ei centrală constă în studiul evoluționist al funcțiilor, trăsăturilor și tendințelor ființelor vii, inclusiv ale omului. Întrebarea ar fi: Cum poate contribui știința viului la înțelegerea sistemelor morale? Ce proceduri, informații, concepte și teorii biologice pot folosi științelor sociale? Ce legătură au problemele eticii cu evoluția organică și ce ar putea fi istoria naturală a sistemelor morale?

După părerea lui M. Bunge<sup>23</sup>, omul reprezintă un sistem socio-bio-psihologic, prezentat până acum în diverse modele reduționiste: *modelul religios*, care privește omul ca o corelație între suflet și corp; *modelul tehnologic*, care gândește omul prin analogie cu computerul; *modelul fizic* în care sunt neglijate creativitatea și spontaneitatea naturii umane, *modelul chimic*, ce concepe omul ca o mașină moleculară genetică; *modelul biologic* consideră omul ca fiind un tip de animal, neglijând specificitatea și proprietățile ce aparțin exclusiv creierului său; *modelul socio-logic* care vede omul drept componentă a societății, ignorându-i dimensiunile biologice. Umanul a mai fost caracterizat și prin dimensiunea sa mentală, ceea ce a facilitat susținerea unui *model cognitivist*. Din unghiul de vedere al filosofiei mentalului *cognitivismul* este o versiune teoretică tare rezultată din apariția științelor cognitive și care susține că orice activitate mentală reprezintă cunoaștere, iar omul este caracterizat de cognițiile sale. Percepția, înțelegerea, învățarea și acțiunea pot fi înțelese doar pe baza modelului cognitiv care cuprinde: emiterea ipotezei, realizarea inferenței, rezolvarea problemei. Datorită unor astfel de idei se consideră că psihologia cognitivă ar fi realizat, pentru înțelegerea omului, ceea ce a realizat Galilei pentru înțelegerea Universului și Darwin pentru înțelegerea biologiei.

\*  
\*   \*

Volumul de studii *Filosofia mentalului. Intenționalitate și experiment*, abordează o temă situată în centrul atenției lumii filosofilor și psihologilor contemporani referindu-se la raportul minte-corp, la experimentul mental, la fundamentele științelor cognitive, toate din perspectiva unei noi ramuri a filosofiei – filosofia mentalului („philosophy of mind” – filosofia minții). Volumul se ocupă de explicarea și înțelegerea naturii „eului”, a actelor sale

cognitive, a stărilor mentale, a conștiinței și comportamentului intențional. Se arată că filosofia mentalului s-a născut ca rezultat al unei cotituri psihologice în filosofie, prin abordarea interacțiunilor psihologiei cu cibernetica și cu Inteligența Artificială, a raporturilor real-psihic, real-mental, logic-psihic, gramatical-mental și a relației computer – self-inteligență – modele mentale.

Volumul se adresează specialiștilor în filosofie, științele cognitive, psihologie și alte științe sociale, dar și studenților și oamenilor de cultură interesați de orientările noi ale spiritualității umane. Studiile sunt semnate de autori străini, (unii protagoniști ai teoriilor noi în domeniu) și de cercetători români. Mare parte dintre autorii străini ai studiilor din volumul de față au contribuit cu sugestii la structurarea volumului. Am obținut de asemenea părerea lui T. Honderich, D. Papineau și T. Crane, cu privire la rolul filosofiei mentalului în contemporaneitate, lucru important având în vedere că ei sunt trei dintre principalii fondatori ai comentariilor și teoriilor asupra naturii mentalului.

Deschidem acest volum cu studiul lui **Ted Honderich**, unul din principalii specialiști în domeniu, profesor de filosofia mentalului la University College din Londra. În numeroasele sale cărți se ocupă de analiza lucrărilor lui Mill, Russell, Davidson și Searle, de aspecte ale determinismului și cauzalității, ale raportului minte-corp, de funcționalism, teoria identității, de anti-individualism. Lucrarea sa principală *A Theory of Determinism: The Mind, Neuroscience, and Life-Hopes* (1988) expune o filosofie deterministă a mentalului, susținând teoria numită *teoria unității*, pe care o consideră mai adecvată ca altele. Cartea examinează consecințele unei astfel de teorii pentru concepțiile asupra libertății, responsabilității, speranței etc. Sunt respinse compatibilismul aparținând lui Hume și continuatorilor săi (teorie ce susține drept compatibile determinismul și libertatea), dar și incompatibilismul tradițional aparținând lui Kant și adepților săi, propunându-se o nouă rezolvare a problemei. Ideile acestei cărți, de 644 pagini, sunt rezumate într-o nouă lucrare *How Free are you? The Determinism Problem* (1993), principala sa preocupare actuală fiind tema naturii conștiinței. „Nu am fost convins de doctrinele funcționalismului, științei cognitive și de altele de același fel, scrie Honderich, așa că încerc ceva nou. Nu va fi exact același lucru cu doctrina lui Brentano, dar va fi ceva asemănător cu o teorie focalizată pe conștiință și nu distrasă de alte lucruri,

cum ar fi de pildă computerul. Conștiința este o chestiune fundamentală deoarece toate aspectele ei menționate mai sus cu referire la teoria determinismului și filosofia mentalului au o relevanță evidentă și în viața practică-politică, socială și personală“.

**David Papineau**, este altul dintre principalii experți englezi (King's College – Londra) în filosofia mentalului inclus în acest volum. El se referă la o tematică foarte dezbătută privitoare la adevăr, referință, explicație, reprezentare, stări mentale, teleologie, survenire, (supervenience), credibilitate (reliabilism), la fizicalism, realism și „moartea“ epistemologiei. Ca principal reprezentant al orientării naturaliste în filosofia științei, orientare realist-fizicalistă originală pe linia pe care se înscriu Lewis, Davidson, Peacocke, McGin, J. Smith, Mc. Fertrige și Lycan; David Papineau susține că ființa umană și capacitățile ei mentale fac parte din lumea naturală, ca atare conștiința poate fi explicată științific, dacă acceptăm că scopul științei nu este exclusiv certitudinea, ci și credibilitatea (reliability). Problemele filosofice care îl preocupă în primul rând, sunt cele despre *reprezentare și cunoaștere* cuprinse într-o teorie *teleologică* a reprezentării și o teorie *reliabilistă* a cunoașterii. „Interesul meu pentru problema reprezentării m-a condus și la alte aspecte ale filosofiei mentalului, în special cele referitoare la *conștiință* și *relația minte-corp* (mind-body), spune el. Interesul pentru tema cunoașterii m-a îndreptat spre elaborarea unei lucrări cu privire la alegerea teoriei în știință și filosofia matematicii. Sunt de asemenea preocupat de concluziile cercetării probabilității, cauzalității, a teoriei deciziei și în această perspectivă de interpretarea mecanicii cuantice. Cred că toate aceste probleme filosofice sunt generate de tensiunile ce apar cu privire la întrebarea: care ar fi cele mai bune teorii pe care le avem despre lume? Oricine este interesat de natura lumii reale va găsi astfel de teme ca fiind importante“.

Despre „ce este filosofia mentalului“ unul din autorii cuprinși în volum, **Tim Crane**, de la University College din Londra, arată că ar fi disciplina ce se ocupă de natura minții și raportul tradițional minte-trup, de intenționalitate (capacitatea mentalului de a se autoorienta spre lucruri), de natura experienței perceptuale din unghiul de vedere epistemologic și metafizic, de cauza mentală, adică de felul cum mintea poate afecta lucrurile din lumea fizică. Scopul său este înțelegerea concepțiilor moderne aspra minții. Eu cred că ele sunt importante în sine, dar și pentru filosofie, notează

Crane, întrucât curentele filosofice contemporane (cele analitice și cele așa-zis continentale) presupun o anumită concepție asupra minții. Pentru a aborda orice problemă filosofică este necesară o anume noțiune a mentalului. Una din dificultățile speciale constă în necesitatea de a ține pasul cu dezvoltarea rapidă a științelor cognitive și încercarea de a surprinde implicațiile lor filosofice. De fapt, dificultățile filosofiei mentalului sunt cele ce apar în cadrul oricărui demers original în filosofie. Se cer examinate critic fizicalismul, percepția, intenționalitatea, cauza mentală, referința, conceptul – ceea ce Crane realizează în cartea sa *The Mechanical Mind* (1995).

În lucrarea de față *Ted Honderich*, se ocupă în studiul său de diferențele între funcționalismul neural și concepția lui Searle despre specificul evenimentelor conștiente; *David Papineau* de teoriile care explică teleologic reprezentarea mentală în termenii scopurilor biologice ale credințelor și ale altor stări mentale; *Tim Crane* se referă la dezbaterile în jurul cauzalității mentale. Dintre specialiștii britanici au mai fost incluși în acest volum și *K.R. Popper*, *J.J. Haldane* și *K.V. Wilkes*. Primul, cu un articol referitor la „abordarea, din perspectiva lumii a III-a, a raportului minte-trup“. După părerea lui *Popper* procesele mentale fiind necesare supraviețuirii biologice a speciei umane, pare mai plauzibilă perspectiva darwinistă (decât cea epifenomenalistă) pentru înțelegerea lumii a III-a, întrucât interpretează rezultatele proceselor mentale ca oferind mijloace de evoluție și de supraviețuire a speciei. *J.J. Haldane* pune în discuție concepția naturalistă, asupra intenționalității a lui *Stalnacker* criticând-o pentru reducționismul său. El caută să demonstreze că legătura dintre stările intenționale și extensionale, între behaviorism și teoria identității, între științele naturii și psihologie exclude teoretic naturalismul monist al lui *Stalnacker*. Pluralismul lui *Aquinas* și *Wittgenstein* i se pare mai adecvat pentru caracterizarea intenționalității ca specific al mentalului.

Abordând raportul minte-trup *K.V. Wilkes* arată că tema a devenit aproape „sfântă“ în filosofia postcarteziană, dar că nu are nici o șansă de a fi rezolvată, ci poate fi cel mult clarificată prin analiza distincției între mental și fizical și plasarea ei în perspectiva funcționalistă.

Spre deosebire de școlile engleze, în SUA filosofia mentalului este concepută ca fundamentare a științelor cognitive în primul rând. Există chiar înfruntări pe această temă între filosofi englezi

Hacker, Honderich, Malcom și cei americani. Din SUA, în volumul de față, am obținut colaborarea lui **Roy Sorensen** care se ocupă de experimentul mental și raportarea lui la legitatea naturii. Studiind fundamentele biologice ale optimismului biologic, Sorensen susține că incompletitudinea teoriei evoluționiste cu privire la natura experimentului mental cere amendamente socio-psihologice și etnometodologice. Tot în secțiunea despre experimentul mental au mai avut amabilitatea să ne ofere materiale **James Brown**, expert canadian în filosofia științei, autor al lucrărilor *Laboratory of the Mind. Thought Experiment in the Natural Science* (1991). Au mai participat cu contribuții la acest capitol specialiștii români Ion Mânzat, Gheorghe Ștefan și G.G. Constandache.

La definitivarea panoramică a prezentării temelor și curenților filosofice cu privire la mental trebuie să adăugăm comentariile din studiul lui **L. Bergström** cu privire la explicațiile pe bază de motive ale acțiunii așa cum apar ele la Hempel și Davidson. Bergström relevă aspectele pozitiviste și hermeneutice care îi diferențiază pe cei doi filosofi în abordarea raportului legi fizice legi psihologice, cauze și motive, explicație și înțelegere. Prezentăm, de asemenea, un articol asupra psihanalizei lui Freud și a implicațiilor ideilor acestuia în analiza stărilor mentale, scris de **G.J. Lock**. Alte aspecte filosofice ale teoriei Inteligenței Artificiale și ale cunoașterii specificului conștiinței morale sunt expuse de Mircea și Mihaela Malița și Liubomira Miroș.

Trebuie să spunem explicit că lucrarea de față s-a orientat spre asocierea unor contribuții ale filosofilor de limbă engleză cu abordări ale specialiștilor români, iar în Bibliografia prezentată sunt cuprinse deliberat numai cărți de limbă engleză, sau traduse în limba engleză. Faptul nu vrea să semnifice lipsa contribuțiilor contemporane în limbile franceză, germană, italiană sau altele, ci doar orientarea cercetărilor noastre de filosofia mentalului spre literatura de limbă engleză și colaborarea directă cu unii dintre filosofi – autori de studii în acest volum – cum ar fi T. Honderich, D. Papineau, T. Crane, K. Wilkes, J. Brown, R. Sorensen, cărora le mulțumim pentru contribuția lor esențială la elaborarea acestei cărți.

## NOTE

<sup>1</sup> J.C. Smith (ed.), *Historical Foundations of Cognitive Science*, Kluwer Acad. Publishers, Londra, 1990.

<sup>2</sup> Simon Blackburn, *Oxford Dictionary of Philosophy*, Oxford Univ. Press, Oxford, 1996

<sup>3</sup> P. Churchland, *The Engine of Reason. The Seat of the Soul* (A Philosophical Journey into the Brain). Mit Press, Cambridge, 1995.

<sup>4</sup> P.M.S. Hacker, *Interviu acordat Angelei Botez pentru Revista de filosofie*, nr. 4-5, 1996.

<sup>5</sup> Vezi J.C. Smith, *op. cit.*

<sup>6</sup> Tim Crane, *The Mechanical Mind*, Penguin Books, Londra, 1995.

<sup>7</sup> David Papineau, *Philosophical Naturalism*, Blackwell, Oxford, 1993.

<sup>8</sup> D. Papineau, *op. cit.*

<sup>9</sup> S. Blackburn, *op. cit.*

<sup>10</sup> Vezi Bibliografie selectivă.

<sup>11</sup> Vezi Bibliografia selectivă.

<sup>12</sup> Vezi Bibliografia selectivă.

<sup>13</sup> Vezi Bibliografia selectivă.

<sup>14</sup> Vezi și Bibliografia selectivă.

<sup>15</sup> J. Habermas – *Moral Consciousness and Communicative Action*, Routledge, Londra, 1989.

<sup>16</sup> P. Ricoeur, *Hermeneutics*, Routledge, 1987.

<sup>17</sup> J. Habermas, *op.cit.*

<sup>18</sup> R. Rorty, *Philosophy and the Mirror of Nature* (1979), *Consequences of Pragmatism* (1982) (vezi Bibliografia selectivă).

<sup>19</sup> Christopher Norris, *Deconstruction: Theory and Practice*, Methuen, Londra 1982.

<sup>20</sup> Chr. J. Lumdsen and E. O. Wilson, *Genes, Mind and Culture. The Coevolutionary Process* (1981).

<sup>21</sup> R. Dawkins, *The Extended Phenotype* (1983), *The Blind Watchmaker* (1987).

<sup>22</sup> R.D. Alexander, *The Biology of Moral Systems*, Aldine de Gruyter, New York, 1987.

<sup>23</sup> M. Bunge, „A Systems Concepts of Society. Beyond Individualism and Holism“, în *Theory and Decision*, 10/1979.



# PARTEA I ORIENTĂRI ȘI CONCEPTE

## CONȘTIINȚĂ, FUNCȚIONALISM NEURAL ȘI SUBIECTIVITATE REALĂ \*

Ted HONDERICH  
(University College, Londra)

### I. CONCEPȚII CU PRIVIRE LA CONȘTIINȚĂ

Funcționalismul și științele cognitive sunt uneori înțelese într-o modalitate filosofică puțin ambițioasă. De exemplu, ele sunt interpretate ca o perspectivă după care dorințele, indiferent de ce se spune despre ele, reprezintă în mod tipic efecte ale unor fenomene percepute și cauze ale comportamentului; sau ca o perspectivă care, căutând să definească dorința fără nici un fel de referire la comportament, se dovedește superficială. Avem aici ceva mai mult decât un truism. Funcționalismul și științele cognitive sunt considerate, diferit și ambițios, ca orientări ce dau adevărata explicație a naturii evenimentelor conștiente proprii sau ale altora. Pe această cale funcționalismul și științele cognitive par a urma behaviorismul dintr-un onorabil trecut al filosofiei mentalului. Se demonstrează, însă, că nu există speranță de adevăr nici pentru ambițioasa afirmație după care o completă explicație a evenimentelor conștiente poate fi dată pe baza susținerii că ele nu conțin altceva decât relații cauzale sau logice, decât input-uri, evenimente interne și output-uri fără să spună nimic despre conținutul relațiilor. O astfel de teză ar conduce la concluzia absurdă că orice fenomen care conține relații normale reprezintă un eveniment conștient. S-au emis multe idei amăgitoare și sunt multe vociferări ostile cu privire la această problemă, dar nimic care să răspundă adecvat reducerii la absurd, produse prin argumentele de mai sus, nu s-a spus<sup>2</sup>.

Afirmația ambițioasă la care ne refeream ne duce și la o a doua absurditate rezultată din obișnuiță, dar nu mai puțin dăună-

---

\* Din Studiul „Consciousness, Natural Functionalism, Real Subjectivity” *American Phil. Quarterly*, vol. 32, nr. 4/1995, p. 366-381.

toare. Este vorba de faptul că identificarea evenimentelor conștiente cu relațiile în sine face imposibilă explicarea realității conștiinței. Realitatea ar fi o problemă de relație și nu de evenimente sau de orice altceva. Deci, concepând conștiința ca substanță-fantomă, funcționalismul și științele cognitive par a o fi redus la și mai puțin decât o relație, chiar realitatea acesteia s-a evaporat.

Insistența asupra faptului că evenimentele conștiinței nu sunt doar relații de genul celor de mai sus, ci au și o altă natură particulară se dovedește consistentă cu o serie de alte afirmații, cum ar fi: propoziția că evenimentele conștiente umane nu sunt decât evenimente neurale, care stau în relație specifică cu un input, cu niște evenimente interne și un output. O astfel de viziune asupra evenimentelor conștiente, despre care știm cel mai mult, este numită *funcționalism neural*. Ea reprezintă o a doua poziție ambițioasă, chiar mai populară decât funcționalismul oficial și științele cognitive, de care ne preocupăm în cele ce urmează. S-a presupus că funcționalismul neural este mai puțin „indigest” decât materialismul eliminativ sau teoria „nimic altceva” decât materialismul, sau decât absoluta negare a conștiinței, susținută de mulți moniști contemporani și de teoreticienii identității creier-minte. În fapt, funcționalismul neural nu pare mai bun decât materialismele amintite, deoarece nimic nu se adaugă la explicația evenimentului conștient dacă îl concepem exclusiv ca relație, chiar un eveniment non-neural, ca de pildă evenimentul silicon, ar putea fi eveniment conștient din moment ce este o relație. Nu putem stabili o proprietate adițională pentru o șurubelniță dacă rolul său poate fi îmbunătățit prin ceva diferit de ea, ca de pildă un deget puternic sau o unghie tare. Ceea ce face ca reducerea evenimentelor conștiente la relații neurale să fie inacceptabilă (...), se pare că este convingerea intimă că evenimentele conștiente sunt diferite de relațiile neurale. Ele au proprietăți neural-cauzale, dar și altele sau numai alte proprietăți. Nu se poate spune care dintre ele nu sunt importante.

Mulți dintre noi avem convingerea că evenimentele conștiente sunt diferite de relațiile neurale într-un mod dat, pentru că apropierea de natura conștiinței celorlalți se face prin amintirea insuficient conceptualizată a evenimentelor conștiinței proprii. În particular, nu putem face nimic pentru a observa gândirea sau simțirea în act ci doar obiectele acestora, ne reamintim de natura gândirii și simțirii imediat după ce actul s-a produs. Insuficienta

conceptualizare a acestei amintiri este analoagă cu obișnuita conștientizare după nuanță a culorii – care vine dintr-o concepție articulată anterior.

O astfel de poziție filosofică ni se pare nesatisfăcătoare pentru că nu face decât să insiste asupra faptului că opinia noastră despre natura conștiinței poate stabili o diferență între conștiință și relațiile neurale. Avem nevoie de o concepție încheată asupra acestei diferențe și, mai mult, de o *concepție asupra naturii conștiinței* tot atât de explicită ca și concepțiile noastre asupra naturii evenimentelor din aria fizicalului, proprietăților și adevărului în general. A nu avea o astfel de concepție se dovedește a fi un lucru nesatisfăcător, faptul neexplicit fiind întotdeauna insuficient pentru filosofie. Nevoia de concepție pune funcționalismul neural într-o poziție bună pentru că el neagă sau pune la îndoială faptul că intuiția noastră asupra conștiinței ar putea stabili o diferență între evenimentele conștiente și relațiile neurale și ne pune în situația de a nu mai fi atât de încrezători în ea<sup>1</sup>.

Prima mea preocupare în această lucrare este să aflu dacă putem avea o concepție asupra naturii conștiinței mai bună decât funcționalismul neural și dacă putem ajunge la o astfel de concepție fără a cădea într-o „apă filosofică tulbure și adâncă”. Atracția funcționalismului neural pentru filosofii minții printre care mă număr și eu constă tocmai în faptul că stă departe de această „apă”, de ceea ce se numește subiectivitate reală sau fundamentală. După cum se întreba Searle, putem oare deschide o perspectivă foarte largă asupra conștiinței compusă numai din adevăruri simple și clare asupra „minții” și bazată pe concepții și doctrine filosofice și psihologice fără nici o obscuritate? În lucrarea sa *The Rediscovery of the Mind*<sup>6</sup>, Searle dă un ghid exemplar de 13 astfel de adevăruri. Ele vor fi expuse în capitolele următoare.

## II. FIZICALITATE ȘI NIVELURI \*

(1) Primul adevăr se referă la faptul că *evenimentele conștiente sunt într-adevăr evenimente fizice*. Eu înțeleg această afir-

---

\* În traducerea de față Cap. II și III au fost prescurtate. Notele 7–21 au fost, de asemenea reduse. S-a efectuat renumerotarea notelor.

mație a lui Searle ca diferită de aceea că evenimentele conștiente ar fi o subclasă a evenimentelor fizice, deci exclusiv relații neurale. Searle are dreptate să considere evenimentele conștiente drept fizicale, în accepția pe care el o dă fizicalului. El crede că evenimentele fizice nu pot fi definite cartezian ca *Res extensa* pentru că electronii, deși sunt fizicali, se definesc ca masă și energie punctiformă. El sugerează că definiția evenimentelor fizicale trebuie dată de domeniul științei actuale.

(2) Găsim diferențe față de funcționalism în concepția lui Searle și atunci când el insistă asupra faptului că evenimentele conștiente sunt evenimente ale creierului cauzate de relații pur neurale prezente în acesta. Al doilea adevăr al său se referă la faptul că *evenimentele conștiente sunt evenimente biologice ale creierului de nivel mai înalt, cauzate de evenimente neurale de nivel mai scăzut, produse de creier*<sup>7</sup>.

### III. SUBIECTIVITATE ȘI UNITATE

(3a) *Evenimentele conștiente sunt subiective*. Ajungem se pare la o altă diferență față de funcționalismul neural, când spunem că evenimentele conștiente fac parte dintr-o categorie de lucruri care au un mod special de existență, o anume ontologie. Orice conștiință există doar drept conștiință a cuiva. Pentru evenimentele conștiente aceasta reprezintă „persoana întâi“, „eul“ de care ele sunt dependente. Eu am o anume relație cu evenimentele mele conștiente. Dependența de persoane a evenimentelor conștiente ne dă primul sens în care ele sunt subiective, mai degrabă decât obiective.

(3b) Se spune că reprezintă o consecință a dependenței de persoană, faptul că un eveniment conștient se dovedește subiectiv în sensul al II-lea, adică nu este accesibil, la fel, pentru toți observatorii. Altcineva nu poate să spună dacă eu mă gândesc la Baden sau nu.

(3c) Faptul că evenimentele conștiente, mai ales cele intenționale sunt dependente de modul de existență al persoanei întâi are și alte consecințe care nu pot fi expuse în termenii subiectivității. De pildă, faptul că informația mea despre lume este perspectivală adică vine dintr-un unghi anume de vedere.

(3d) În fine, raportarea la modul de existență al persoanei întâi implică sau are drept consecință și faptul că evenimentul conștient poate fi subiectiv și în al III-lea sens, adică persoana îl vede asemănător cu ceva sau doar crede că îl vede asemănător cu ceva.

(4) S-a spus despre *conștiință* că *ar include două feluri de unitate*: una existând la un anumit timp, alta deasupra perioadelor de timp. Experiența a trei obiecte percepute la un moment dat le vede ca părți ale aceluiași eveniment conștient. Aceasta este caracteristica conștiinței de care vorbește Kant în termenii unității transcendente apercceptive. Distinctă față de o astfel de unitate „verticală” este unitatea „orizontală”.

Conștiința mea cu privire la începutul apariției propozițiilor gândite sau spuse continuă și după ce începutul a trecut. Această unitate poate include nu numai memoria iconică menționată, ci și memoria pe termen scurt.

(5) Evenimentele conștiente *pot fi clasificate în cinci tipuri*, în legătură cu cele cinci simțuri ale omului.

(6) Ele se pot distinge și datorită fundamentului lor structural, după cum arată psihologia gestaltistă, incluzând:

(7) familiarizare atotcuprinzătoare.

(8) revărsare-deversare, gândurile se separă de, sau se conectează cu alte gânduri.

(9) centrul și periferia sunt determinate de atenția selectivă și de nivelele atenției.

(10) condiții limitative privind localizarea spațio-temporală-socio-biologică pentru mine a evenimentelor mele conștiente.

(11) stări emoționale.

(12) dimensiunea plăcere-neplăcere.

În fiecare din aceste sensuri funcționaliștii neurali găsesc factori care să se potrivească concepției lor despre minte. De aceea noi nu am găsit ceea ce ne interesează, poate, în cele ce urmează.

#### IV. INTENȚIONALITATEA

(13) Searle distinge concepțiile „ersatz“ ale intenționalității de cele care abordează lumi reale și le îndepărtează efectiv pe cele de mai sus.

El este de acord că intenționalitatea (adică reprezentativitatea, calitatea de a fi despre ceva, a avea semnificație sau orientare) conținutului evenimentelor conștiinței nu poate fi descrisă ca simplă relație cauzală între ele și lucrurile lumii, incluzând relațiile de tipul celor propuse de externaliști sau chiar cu un conținut vid. Încercările de a naturaliza intenționalitatea prin reducerea ei la factorii neurali, care în relația cauzală devin efecte și probabil și factori de mediu extern<sup>8</sup> lucrând în alt mod, se dovedesc nereușite. Dar ce se înțelege prin faptul că un eveniment ar fi intențional?

Mulți filosofi consideră că o bună explicație a conținutului intențional poate ajuta mai mult ca orice la caracterizarea evenimentelor conștiente. Searle face parte dintre aceștia, în profida unor idei discordante ale sale la care vom ajunge. După părerea proprie, el nu se include între cei ce susțin că problema conștiinței nu reprezintă altceva decât problema intenționalității. El spune, la un moment dat, că încercarea de a naturaliza conținutul intențional ajunge să elimine subiectivitatea acesteia<sup>9</sup>, dar, nu este clar care aspect particular al subiectivității se pune în discuție.

Concepția asupra intenționalității conținutului, care susține faptul că fiecare eveniment conștient este subiectiv în sensul de a fi dependent de o persoană nu poate propune o interpretare a evenimentelor conștiente, diferită de concepțiile „ersatz“. Situația este aceeași cu cea observată mai devreme în legătură cu dependența de persoană, folosită direct (distinct de cale intenționalității) în speranța de a găsi o explicație a evenimentelor conștiente<sup>10</sup>.

Putem oare interpreta intenționalitatea conținutului prin arătarea faptului că fiecare eveniment conștient este subiectiv în sensul al doilea, de a nu fi egal accesibil oricărui observator? Nici această perspectivă epistemologică nu pare a găsi modul de

caracterizare a naturii intenționalității. Sau, putem caracteriza intenționalitatea în mod fundamental, prin relevarea faptului că evenimentele conștiente sunt perspective sau au nuanțe perspective? Răspunsul e negativ. Factorul în chestiune este familiar și acceptat de toți, dar din punct de vedere al teoriei intenționalității este pre-teoretic și îl folosește orice interpretare. Ajung oare mai departe pe calea propoziției, (3d) evenimentele conștiente au aspectul a ceva ce se aseamănă sau simt că se aseamănă? Răspunsul este tot nu pentru diferite motive, unele arătate mai sus, când subiectivitatea în al III-lea sens a fost privită cu speranța că va duce direct la o concepție distinctivă asupra evenimentelor conștiente. Altă explicație a fost menționată în legătură cu (3a) și (3b). Reprezintă un fapt interpretarea evenimentelor conștiente ca posedând subiectivitatea, deși nu toate sunt intenționale. Deci, nu se poate explica intenționalitatea prin referire la subiectivitatea în chestiune.

Proiectul caracterizării intenționalității în alt mod decât cel al concepțiilor „ersatz“ nu pare să înainteze prea mult chiar când e vorba de o doctrină normativă a intenționalității: noțiunile intenționale ar stabili standarde ale adevărului, raționalității, consistenței etc.<sup>11</sup>. Chiar dacă presupunem că doctrina normativității este acceptabilă, ea nu ne folosește în acest caz, deoarece faptele cu privire la intenționalitate suferă de aceeași lipsă de explicație. Același fel de dezamăgire produce și doctrina cu privire la faptul că orice conținut intențional depinde de ceea ce se numește *background* identificat cu anumite capacități, abilități și *know how* general<sup>12</sup>.

În acest punct trebuie să ne referim la discordanța în concepția lui Searle pe care o anticipasem. Nu putem studia intenționalitatea conținutului fără să studiem conștiința, intenționalitatea nu poate fi înțeleasă decât în relație cu conștiința de care este dependentă<sup>13</sup>. Recomandarea răstoarnă ordinea cercetării pe care tocmai o contemplasem, ea merge la concepția despre conținut intențional de la o concepție deja existentă a conștiinței, adică invers față de strategia pe care o observasem. *Explanans* și *explanandum* sunt inversate. Trebuie oare să alegem între cele două strategii? Cred că putem fi mai puțin tradiționali și unidirecționali, acceptând, fără să ne temem de circularitate, dubla mișcare; (a) într-un fel,

subiectul intenționalității este luminat de subiectul conștiinței și (b) în altul subiectul conștiinței este luminat de subiectul intenționalității.

Dar dacă pornim pe această cale, cercetând intenționalitatea prin întoarcerea spre conștiință, vom constata o dificultate supărătoare în lipsa unei concepții asupra conștiinței. Mai mult, dacă mergem la propozițiile (1, 2, 4, 12) despre conștiință, care nu au fost folosite până acum pentru a clarifica problema intenționalității în modul dorit, constatăm că nici ele nu ne pot fi de folos. Las acest exercițiu pentru cititor.

Să ne întorcem la (b) încercând să clarificăm tema conștiinței, dând atenție intenționalității. Nu putem realiza însă acest lucru atâta timp cât ne lipsește o bună interpretare a intenționalității, care nu a putut fi găsită în conexiunea între intenționalitate și subiectivitate. Ceea ce trebuie luat în considerare este demersul asupra intenționalității conținutului mental, realizat de Searle în cartea sa *Intentionality*, în care intenționalitatea derivă din limbaj.

Această abordare consideră evenimentul conștient intențional ca depinzând de conținutul (I) care fixează (II) o condiție a adevărului sau o „condiție a satisfacției“ sau (III) de un model psiho-logic potrivit căruia evenimentul reprezintă o credință, o dorință sau altceva și care determină (IV) direcția de adecvare. În cazul unei credințe, ea trebuie să se potrivească cu lumea mai degrabă decât lumea să fie schimbată pentru a se adecva credinței. Iată o interpretare ce pare a fi cea mai bună dintre cele de același tip. Dar de care tip? Acela care se referă la ceea ce se numește „setul de proprietăți logice ale evenimentelor conștiente intenționale“. Este greu să rezisti încercării de a defini astfel natura fundamentală a intenționalității. Poate aceasta reprezintă expresia cea mai bună a neclarității cu privire la ce ar conține explicația lui Searle în dezacord și inconsistent cu funcționalismul neural sau cu citirea lui naturalizată. La Searle pare a exista un factor de dezacord, dar care ar fi el? Ar putea să fie faptul că el nu admite un anumit „statut ontologic“ al evenimentelor în chestiune. Searle o spune explicit. Acest statut apare când se susține ca (2) evenimentele conștiinței sunt evenimente biologice de înalt nivel ale creierului cauzate de evenimente neurale de nivel inferior.

Totuși, eu nu doresc să forțez nota pe această linie până nu e nevoie de așa ceva. Și nu e nevoie pentru abordarea noastră să



realizăm caracterizarea naturii evenimentului conștient, pentru că, după cum am remarcat și este larg acceptat, nu toate evenimentele conștiente sunt și intenționale, fapt foarte adesea lăsat la o parte în filosofia contemporană a mentalului. Pentru noi e important de știut că avem o slabă speranță să putem explica evenimentul mental pe baza proprietății intenționalității și a reflecției asupra acesteia. Există evenimente conștiente paradigmatic care uneori nu au proprietatea intenționalității, cum ar fi senzațiile și stările emoționale<sup>16</sup>.

Este adevărat că enumerarea de către Searle a celor treisprezece caracteristici, care se constituie ca un grup structurat de trăsături ale conștiinței și tot ceea ce s-a spus despre acestea duce la creșterea încrederii că ne apropiem de natura conștiinței. În parte, pentru că include și reamintiri bune, procedura întărește și convingerea cu privire la posibilitatea cunoașterii naturii conștiinței, ceea ce nu este puțin.

Dar să stabilim principala concluzie a acestui studiu: Reflecțiile și adevărurile anumitor curente și doctrine filosofice și psihologice cu privire la mental nu ne oferă, despre natura conștiinței, o concepție distinctă de funcționalismul neural<sup>17</sup>. Mai mult, toate afirmațiile se dovedesc consistente cu funcționalismul, sau ca în cazul (3d), lipsite de utilitate. Nu avem deci o bază suficient de conceptualizată sau o premiză majoră pentru a argumenta împotriva funcționalismului neural. Nici unul din cele 13 puncte nu ne dă viziunea așteptată pentru aceasta. Searle își rezumă poziția față de natura conștiinței la afirmația că ea apare la un nivel biologic ridicat cu tot ce are ca subiectivitate, intenționalitate etc. și este cauzată de evenimente neurale<sup>18</sup>. Nu avem, deci, prin cele treisprezece trăsături o nouă concepție distinctă asupra conștiinței.

Searle și filosofii de aceeași orientare nu sunt ca acel om beat, pe care el îl menționează comparativ cu funcționalistul și specialistul în științe cognitive, care și-a pierdut cheile de la mașină într-o zonă întunecată de tufișuri, dar le caută într-un luminiș pentru că acolo vede mai bine. Searle și filosofii ca el caută în locul potrivit, dar fiind preocupați să-i arate omului beat tufișurile și fiind, totuși, atrași spre strada luminată, lasă cheile în umbră.

## V. SUBIECTIVITATEA REALĂ

Searle, care încercase mai înainte să abordeze conștiința prin mijloacele indirecte ale reflecției asupra intenționalității limbajului<sup>19</sup>, în lucrarea *The Rediscovery of the Mind* urmează calea așa numitului *realism mental*. El se ocupă direct de realitatea evenimentelor conștiente mai degrabă decât să caute factori divergenți mai accesibili care să-l ducă până la urmă, ocolit, spre aceste evenimente. Funcționalismul și știința cognitivă sunt prime exemple pentru astfel de căi ocolite. Un alt exemplu notabil îl reprezintă încercarea de a specifica și ce este conștiința cu ajutorul anumitor proprietăți logico-lingvistice ale acelei părți a limbajului prin care ea se exprimă<sup>20</sup>.

Dar o concepție distinctă asupra conștiinței va putea fi produsă și de experiența noastră, de înțelegerea obișnuită a conștiinței pe care o posedăm. Ea va avea o mai mare extindere decât reflecțiile lui Searle, reușind să nu sucumbă prea repede în fața cererilor de literalitate, claritate și explicitare și să nu piardă din vedere realitatea asupra căreia are deschidere, respectând-o în relatările asupra tipurilor de subiectivitate menționate mai sus: unitatea, intenționalitatea și a.m.d.

Această concepție nu ne scoate din mister, dar încearcă să-l disipeze. Ea nu ocolește „apa adâncă și tulbure“, dar caută să extragă și ceva clar din ea. Cred că astfel de concepții despre conștiință fac parte din aceeași familie cu concepția lui Brentano, care vorbea despre conștiință ca de o activitate *referitoare* la un conținut sau de o activitate direcționată spre un obiect<sup>21</sup>. De asemenea, acest alt fel de concepție va căuta să scape de metaforă, de obscuritatea și neexplicitarea lui Brentano și a concepțiilor de același tip. Ea ocolește și dezastrul adus de acceptarea unui „homuncul“ în sinele persoanei, adică de a tolera misterul în centrul istoriei sau construcției conștiinței, reușind, în final, nu numai să dea o viziune suficient conceptualizată dar și una literală, clară, explicită și coerentă, în locul surprinderii intuitive a conștiinței.

Dacă încercăm să construim o concepție distinctă și coerentă asupra evenimentului conștient putem începe cu ideea că fiecare din aceste evenimente are două proprietăți: una din ele fiind

conținutul, care se cere interpretat în așa fel încât să se evite fenomenalismul și să se prezerve faptul că lumea este dată. Nu este ceva subiectiv să fiu conștient de un fapt normal. Cu referire la cealaltă proprietate, sau parte a evenimentului conștient, trebuie să vorbim despre subiect sau despre subiectul-parte, care nu este, evident, o persoană, în sensul corpului persistent și/sau a unei secvențe de relații interne, și nici o persoană, eu sau ego în sensul cartezian, ori în alt sens. O astfel de concepție promite să includă o abordare a ceea ce pare a fi subiectivitate reală sau fundamentală, distinctă de subiectivitatea legată de dependența personală, de accesul inegal, de ceva ce este asemănător sau este simțit ca asemănător.

Relația între cele două proprietăți sau părți ale evenimentelor conștiinței, pare a fi una de interdependență. Pentru numeroase rațiuni, una dintre ele, aceea de a evita fenomenalismul, adică, introducerea obiectelor subiective în conștiință, relația nu va putea fi descrisă drept conștiință a conținutului. Ea este, de asemenea, o relație deplin distinctă de cele cu care ne-am ocupat până acum. Este o relație care intră în multe, dar nu în toate evenimentele conștiente, ca relație de intenționalitate între conținut și obiect<sup>22</sup>. Este dificil de evitat supoziția că relația a doua va fi luminată până la urmă tot de prima.

Eu nu am reușit să pot oferi până acum o concepție distinctă și adevărată asupra evenimentului conștient, cu toate eforturile depuse să ajung la una<sup>23</sup>. Lipsa ei nu mă împiedică, însă, să cred că acesta este scopul de urmărit, înainte de toate, în *filosofia minții*.

S-au spus destule despre o concepție distinctă și adevărată asupra conștiinței pentru a putea trage o a doua concluzie principală. Ne întrebăm dacă, după toate cele implicate în concepția nouă a evenimentelor conștiente, este posibil ca teoria să fie doar o anticipare sau deschidere spre funcționalism, adică să fie consistentă cu funcționalismul neural? Va fi ea același lucru cu cele treisprezece trăsături ale conștiinței, deja stabilite? Răspunsul este negativ.

Răspunsul pentru care spun aceasta este faptul că funcționalismul neural nu reprezintă o concepție asupra naturii evenimentului conștient în sine. Faptul că evenimentul apare drept conștient rămâne în „cutia neagră” deoarece trăsătura evenimentului conștient de a fi neural se referă doar la o descriere a relațiilor sale externe.

Concepția adecvată va fi fundamental relaționistă, chiar dacă va fi vorba de o relație internalistă. Va fi ea doar o parte a noului funcționalism? Se poate replica: nu. Relațiile de care se preocupă în principal funcționalismul neural și care îl definesc sunt între conținutul evenimentelor conștiinței și cunoscutele: input, conținutul altor evenimente conștiente, output. La fel se prezintă și alte funcționalisme, precum cele de care am discutat mai sus.

Și aceasta nu este tot. Dacă relația între proprietatea-conținut și proprietatea-subiect a evenimentului conștient este nomică și logică nu mai apare necesitatea de a vorbi despre funcționalism, unicul caracter al evenimentului conștient fiind relația nomică și logică<sup>24</sup>. Cercetarea acestui caracter va sta departe de sursele oricărui funcționalism, analog cu mașina lui Turing și succesorii acesteia, de istoria și impulsurile funcționaliste. Nu ar fi exagerat să observăm că puterea doctrinei funcționaliste constă în evitarea misterului și a „apei adânci”, cu care o concepție distinctă adecvată conștiinței trebuie să se înfrunte.

## **VI. DUALISM ȘI MONISM. EXTREMISME.**

Două teme rămân: prima, referitoare la problema relației între conștiință și creier, este problema minte-trup. Aceasta însă nu va fi principala noastră temă pentru că ne preocupă mai ales natura conștiinței. Dar cu siguranță în cadrul abordării ei, există răspunsuri și la chestiunea relației minte-trup, pentru că este normal ca relațiile asupra conștiinței să cuprindă astfel de răspunsuri. În particular, însă, noi considerăm greșită orice abordare a conștiinței care conține așa – numitul dualism.

Searle notează că termenii „dualism”, „monism” sau „materialism” și „teoria identității” precum și variantele lor nu mai sunt azi de mare utilitate<sup>25</sup>. Un motiv ar fi acela că orice monism are proprietăți dualiste și orice dualism nu este mai dualist decât un monism tipic.

Searle insistă asupra faptului că însăși concepția sa nu trebuie caracterizată drept materialism sau dualism. Ce putem spune

despre insistența aceasta? Desigur, Searle are dreptate să obiecteze la denumirile de „materialism“ sau „dualism de proprietăți“ date concepției sale întrucât ele presupun separația prin conținut a celor două domenii – conștiința și fizicul, ca și cum nu ar exista nimic comun între ele.

Reluând teza (1) după care evenimentele conștiinței sunt fizicale, adică domeniul conștiinței se include în fizical, Searle arată că „dualismul proprietăților“ presupune o susținere implicită a doctrinei introspecției, a accesului privilegiat sau a observației interioare pentru el de neacceptat. El rezistă acestei denumiri și pentru că indică concepții ce susțin existența în câmpul de forță mai degrabă a „substanțelor“ decât a particulelor sau a faptului că orice conștiință reprezintă o „substanță“ într-un sens particular, sau o altă idee pe care Searle o respinge: insolubilitatea problemei mind-body<sup>26</sup>. În concluzie, el ar putea avea dreptate când afirmă că propoziția (2) cu privire la cele două nivele ale evenimentelor nu-l face dualist. Ce contează dacă, specificul sau natura evenimentelor la cele două niveluri sunt diferite. Toate aceste insistențe ale lui Searle se dovedesc de mare valoare având în vedere tendința filosofiei contemporane a minții de a conduce spre argumente indefinite. Dar nici una nu duce dincolo de alte două lucruri adevărate. Searle se declară împotriva reducționismului și de aceea insistă și asupra faptului că evenimentele noastre conștiente au și proprietăți diferite de cele pe care le recunoaște funcționalismul neural<sup>27</sup>. Opinia aceasta răstoarnă însă ceea ce el afirmase mai sus și ne dă dreptul de a considera concepția sa drept un „dualism al proprietăților“. Concluzia noastră nu va fi afectată de faptul că vorbind de astfel de probleme, Searle urmează o cale mai apropiată de ceea ce se preferă de obicei și astfel salvează evenimentele. Rațiunea pentru care vorbim despre „dualism al proprietăților“ poate fi pur și simplu faptul că există proprietăți sau trăsături neural-cauzale ale creierului și proprietăți sau trăsături care cer altă caracterizare.

Dar, mai trebuie spus și altceva. Este un lucru să insiști asupra faptului că evenimentele conștiente în sine nu sunt numai relații neurale și altceva să oferi o concepție distinctă asupra lor. Cercetarea noastră arată că Searle nu a reușit să realizeze o astfel de viziune. Nici una din cele treisprezece trăsături și nici suma lor nu au reușit așa ceva. Astfel că poziția sa reală se dovedește consistentă cu funcționalismul neural.

Concluzia a treia pe care o tragem se referă la faptul că viziunea ce reiese după o examinare atentă nu este dualistă doar în sensul în care a fost anunțată. Există diferență între intenție și realizare. De aici rezultă cea mai bună explicație a existenței unor răspunsuri conflictuale la Searle.

Concluzia a patra privește faptul că acea concepție distinctă și adecvată a conștiinței despre care vorbeam mai sus ar putea fi, în intenție și chiar în fapt, corect descriptibilă ca reprezentând un dualism al proprietăților, ceea ce nu o face însă diferită de teoriile contemporane ale identității incluzând monismul anomalistic a lui Davidson. Ca și aceste teorii, într-un anume sens ea poate fi corect descriptibilă și ca „monistă”<sup>28</sup>. Cel mai important este faptul că o astfel de concepție nu se dovedește de neacceptat și nu stă printre teoriile fără viitor.

O ultimă observație ne face să întrebăm: Este oare prospectul unei astfel de concepții asupra conștiinței, dacă nu de neacceptat, măcar excesiv? Există mulți autori care răspund pozitiv și eu admit că înțeleg de ce o fac. Acest proiect de concepție susține că există două feluri de evenimente care apar în capetele noastre: unele care sunt evenimente fizicale în sensul obișnuit și care implică spațialitate. Sau, altfel spus, sunt două feluri de proprietăți fizicale ale minții noastre: un fel de evenimente sau proprietăți care se referă numai la neuroni și relații cauzale fundamentale, specifice funcționalismului neural. Celălalt fel de evenimente, cele ale conștiinței, cuprind ceea ce s-a numit de la început evenimente parțial subiective sau – parțial substanțiale. Trebuie admis că astfel de viziune<sup>29</sup> nu s-a dovedit a avea de a face cu materia conștientă, dar nici cu travestiri ale ei.

Nu încerc să dovedesc incorectitudinea lor, dar țin să fac câteva remarci. Multe din ele rezultă din așteptarea unor lucruri pe care Searle ar fi trebuit să le spună pe baza concepției conștiinței pe care a promis-o, dar nu a realizat-o.

Prima remarcă se referă la faptul că o astfel de concepție apare ca adevărată experienței noastre, surprinderii de către noi a conștiinței, adevărată față de ceea ce știm fără să ne îndepărtăm de aprehensiunea subiectului. O remarcă secundă ar fi că ea depinde de doctrina introspecției prin faptul indubitabil al memoriei. A treia remarcă se referă la faptul că trebuie să ne fie clar că evenimentele și proprietățile mentale la care ne referim

sunt fizicale în sensul obișnuit. Ele nu constituie o altă lume. A patra remarcă corelată cu a treia se referă la faptul că nimic nu stă în calea includerii lor în contextul explicativ al conexiunilor nomice și cvasilogice. Mai concret, nimic nu stă în calea faptului de a deveni corelate nomice ale evenimentelor neurale sau cauze și, de asemenea, efecte ale lor. A cincea remarcă și ultima spune că trebuie observat, deși pare absurd că o astfel de teorie asupra evenimentelor conștiente se dovedește inconsistentă cu concepția științifică despre lume. Neuroștiința este de mult timp angajată în studiul conexiunilor nomice între un anumit tip de evenimente neurale și evenimentele conștiente și între evenimente neurale și evenimente despre care avem întotdeauna o părere și despre care sperăm să găsim și o concepție distinctă și adevărată.

## NOTE

<sup>1</sup> Mulțumiri pentru comentariile făcute asupra primelor forme ale acestui studiu lui Finn Collin, Tim Crane, Owen Flanagan, John Heil, O.R. Jones, E.Y. Lowe, Simon Mathew, Gregory McCulloch, Jane O Grady, Ingmar Persson, G. Segal, M. Targett și referenților anonimi A.P.Q. Nu suntem însă în perfect acord cu ei.

<sup>2</sup> Cea mai cunoscută și mai bună formă a acestui argument este argumentul *camerei chinezești* a lui J. Searle, în „Mind, Brains and Programs” în *Behavioral and Brain Sciences*, vol. 3 (1980) pp. 417–24. O salut ca fiind o contribuție de semnalare.

<sup>3</sup> Donald Davidson, „Mental Events” în *Essays on Actions and Events* (Oxford: Clarendon, 1980), p. 214. Pentru reflecții asupra fertilității denumirii de „teoria identității”, vezi ultima secțiune a acestui studiu.

<sup>4</sup> Pentru argumente asupra inacceptabilității, rebarbativității și incoerenței funcționalismului neural, vezi lucrarea mea: „Functionalism, Identity Theories, the Union Theory”, în Szubka și Warner (eds.). *The Mind-Body Problem: The Current State of Debate* (Oxford, Blackwell, 1994).

<sup>5</sup> Cred că suntem pe drumul cel bun și într-o poziție filosofică corectă de respingere a funcționalismului neural dacă putem da o prezentare explicită cel puțin unei trăsături a conștiinței care nu poate fi acordată cu funcționalismul neural. Nu vom putea însă ajunge la aceasta restrângându-ne la adevăruri clare și simple despre mental ci prin idei filosofice și psihologice. Una dintre ele este de mare interes în această conexiune.

<sup>6</sup> John Searle, *The Rediscovery of the Mind* (London, Cambridge, MA MIT Press 1992).

<sup>7</sup> Vezi cartea mea *A Theory of Determinism. The Mind, Neuroscience, and Life Hopes*, (Oxford, Clarendon, 1988) sau *Mind and Brain* (Oxford, Clarendon, 1990), p. 87–89.

<sup>8</sup> Viziunea mea critică asupra externalismului este dată în „The Union Theory and Anti-Individualism”, în J. Heil and A. Mele (eds) *Mental Causation*, (Oxford, Oxford Univ. Press, 1993).

<sup>9</sup> p. 50.

<sup>10</sup> Vezi secțiunea a III-a.

<sup>11</sup> pp. 51, 238.

<sup>12</sup> cap. 8.

<sup>13</sup> pp. 18, 84, 227.

<sup>14</sup> p. 83

<sup>15</sup> *Intentionality. An Essay in the Philosophy of Mind* (Cambridge: Cambridge: Univ. Press 1983).

<sup>16</sup> Câțiva filosofi înclină să creadă că ceilalți au o idee greșită despre intenționalitate sau că ei confundă această relație cu o alta, internă evenimentului conștient, despre care ceva va fi spus mai jos: Din întâmplare părerea că nu poate fi elucidată conștiința prin intenționalitate pentru că nu întreaga conștiință este intențională se dovedește diferită de părerea inițială, că intenționalitatea nu poate fi explicată prin subiectivitate pentru că întreaga conștiință este subiectivă, dar nu toată este intențională.

<sup>17</sup> Cred că argumentul *camerei chinezești* și alte argumente din *The Rediscovery of the Mind* (cap. 9), cu toată admirația mea pentru ele, se cer întărite printr-o teorie pozitivă asupra naturii conștiinței.

<sup>18</sup> Această concepție asupra relației minte-trup mi se pare deschisă la numeroase întrebări. Una se pune chiar împotriva bunelor intenții ale lui Searle din cauza cursului actual epifenomenalist al filosofiei minții. Vezi, *A Theory of Determinism sau Mind and Brain*, pp. 99–102.

<sup>19</sup> *Intentionality*.

<sup>20</sup> R.M. Chisholm, *Perceiving: A Philosophical Study* (Ithaca, Cornell Univ. Press, 1957), Davidson, „Mental Events”, *op.cit.*

<sup>21</sup> „The Distinction Between Mental and Physical Phenomena” în *Psychology from Empirical Stand Point*, ed. O. Kraus and Mc Alister (Londra, Routledge and Kegan Paul, 1973).

<sup>22</sup> În legătură cu cele două relații el nu vorbește totuși explicit decât de prima. Searle scrie: „un eveniment” nu poate fi doar conștient pentru că trebuie să răspunzi și la întrebarea: De ce anume este conștient? Dar, „conștient de” nu este totuna cu „intenția de” (p. 84, cf. pp. 130–131). Există relația între subiect și obiect (intenționalitatea). Fiecare eveniment conștient conține prima formă de relație (de „of”). Multe, dar nu toate evenimentele conștiente conțin și a doua relație („of”). Cu prima relație Searle se află foarte aproape de chestiunea esențială a subiectivității reale.

<sup>23</sup> „Seeing Things” în *Synthese*, 1994.

<sup>24</sup> Un punct similar apare despre intenționalitate: Care alte relații decât cele nomice pot apărea între conținut și obiect? Problema se pune în legătură cu caracterul acestor relații nomice. Aserțiunea relației fără conținut nu ajută.



<sup>25</sup> pp 2, 4, 14–15, 25–26, 54–55, cf. lucrarea mea, „The Union Theory and Anti-Individualism“.

<sup>26</sup> Obiecțiile lui (pp. 100–105) la argumentul insolubilității nu mi se par eficiente, cum ar trebui să fie. Un argument, atribuit lui Nagel, susține că dacă materia explică conștiința atunci trebuie să fie o conexiune necesară între cele două, o conexiune de fapt, nu o conexiune conceptuală. Obiecția se referă la faptul că explicația nu depinde de conexiunea conceptuală, ci de cea nomică. Conexiunea cauzală nu este conceptuală. Al doilea argument, pentru insolubilitate, derivat din Nagel și atribuit lui Colin Mc Ginn, este acela că orice conștiință este o substanță de care suntem conștienți prin introspecție, iar creierul este, prin contrast, ceva de care devenim conștienți prin percepție. S-ar putea să nu avem posibilitate de a fi conștienți de o legătură explicativă, pentru că nu există al treilea tip de conștiință... Nu vom putea avea nici un fel de explicație dacă fiecare explicație va trebui aplicată în sensul dat. Vezi, *A Theory of Determinism, sau Mind and Brain*, cap. I.

<sup>27</sup> Este fără îndoială că Searle se opune reducționismului exemplificat de funcționalismul neural, așa cum se vede și din cartea sa (vezi paginile XII, 28, 112, 116, 118, 169–199). Totuși, spusele sale că „mentalul este neurofiziologic la cel mai înalt nivel (nota 3 p. 253 cf. 161) este cel puțin fără sens în acest context. El nu merge mai departe pe linia antireducționistă după care „neurofiziologia“ trebuie să admită diferite nivele de descriere, dar nici unul dintre aceste nivele obiective nu este nivelul subiectivității“ (p. 169). Deși consideră subiectivitatea, după cum am văzut, ca fundamentală pentru conștiință o plasează în sfera fizicalului.

<sup>28</sup> Una din aceste teorii este monismul anomalistic al lui Davidson. Vezi, „Mental Events“.

<sup>29</sup> O astfel de viziune este teoria unificării conținând conexiuni nomice între conștiință și evenimentele sau proprietățile neurale, prezentată în lucrarea mea, *A Theory of Determinism și în Mind and Brain*.

Traducere Angela BOTEZ

# ADEVĂR ȘI TELEOLOGIE \*

David PAPINEAU

(Universitatea King's College, Londra)

## 1. INTRODUCERE

O seamă de autori recentî au susținut că reprezentările mentale ar trebui explicate teleologic, în termenii scopurilor biologice ale opiniilor și celorlalte stări mentale<sup>1</sup>. O idee ceva mai veche este aceea potrivit căreia condiția de adevăr a unei opinii este acea condiție care garantează că acțiunile bazate pe opinia respectivă vor reuși.<sup>2</sup>

În articolul de față doresc să arăt că aceste două idei sunt complementare. Teoria teleologică este inadecvată dacă nu încorporează și teza că adevărul este garanția acțiunii reușite. Invers, abordarea condițiilor de adevăr prin prisma garanției succesului este incompletă, dacă nu e plasată într-un context teleologic.

Voi continua în felul următor. În secțiunea ce urmează voi explica de ce reprezentarea mentală este problematică din punct de vedere filosofic. Voi arăta apoi, în Secțiunea 3, care este rolul jucat de noțiunile reprezentationale în explicația acțiunii. Vom fi astfel conduși către o versiune a abordării condițiilor de adevăr prin garanția reușitei, iar în Secțiunea 4 voi elabora și apăra această abordare. Secțiunea 5 va arăta apoi de ce abordarea condițiilor de adevăr prin prisma garanției succesului trebuie să fie încorporată într-o teorie mai generală a reprezentării mentale. În secțiunea finală voi arăta cum se poate răspunde unei obiecții standard la adresa teoriei teleologice.

## 2. REPREZENTAREA MENTALĂ

De ce ar trebui să considerăm că reprezentarea mentală este problematică din punct de vedere filosofic? Un răspuns sumar ar

---

\* „Truth and Teleology“, în *Acta analytica. Philosophy and Psychology*, nr. 6, 1990, pp. 113–133.

fi că orice tip de reprezentare este problematică sub aspect filosofic. Nici o metafizică rezonabilă nu ar trebui să susțină că relațiile semantice ar fi trăsături primitive ale lumii. Relațiile semantice trebuie să fie cumva explicate în termeni non-semantici.<sup>3</sup>

Așadar, reprezentarea mentală este problematică, inițial, în calitatea ei de caz particular al unei probleme *generale*: cum este posibil ca ceva – inclusiv stările mentale – să stea pentru lucruri diferite de el însuși? Legat de reprezentarea mentală există însă și o dificultate mai specifică. Cel mai obișnuit rol pe care îl joacă în gândirea noastră cotidiană opiniile, dorințele și alte stări mentale reprezentationale este cel deținut în explicația acțiunii. Dar ce anume are de-a face explicarea acțiunii cu reprezentarea? Cum se face că explanatorii acțiunii reprezintă alte lucruri decât sunt ei înșiși? Explanatorii accelerației – forțele – nu reprezintă și alte lucruri în afara lor înșiși. Și atunci, de ce ar face-o explanatorii acțiunii?

O strategie naturală ar fi să se încerce soluționarea problemei generale a reprezentării mentale răspunzându-se la această întrebare mai specifică despre legătura dintre reprezentarea mentală și explicarea acțiunii. În definitiv, atunci când în explicația acțiunilor invocăm opinii și dorințe, le identificăm întotdeauna reprezentational, drept opinia că *cutare-și-cutare* sau dorința de *cutare-și-cutare*. Pare astfel credibil că forțele reprezentationale ale stărilor mentale sunt cumva importante pentru semnificația lor explicativă. Poate că o înțelegere detaliată a felului cum acționează acestea ne va îngădui să înțelegem cum anume își află locul reprezentarea mentală într-o lume intrinsec non-representatională.

Cu toate acestea, există un motiv *prima facie* de a ne îndoi că strategia menționată ar putea să funcționeze. Haideți să ne concentrăm asupra opiniilor. Aspectul reprezentational al opiniilor poate fi judecat în termenii condițiilor de adevăr: orice opinie dată va fi adevărată dacă lumea este aidoma cu ceea ce afirmă ea – altminteri, ea va fi falsă. Necazul, acum, este că această diferență între adevăr și falsitate pare irelevantă în raport cu semnificația explicativă a opiniilor. Asta se întâmplă pentru că, desigur, atunci când explicăm o acțiune oarecare în termenii unei oarecare opinii, tot ce contează este ca agentul *să fi avut* acea

opinie, și nu faptul că ea e adevărată sau falsă: o dată ce ai o opinie, ea va avea aceeași influență asupra ta, indiferent dacă e adevărată sau complet greșită.

Permiteți-mi să formulez principiul ce stă la baza acestui tip de judecată:

(A) Adevărul sau falsitatea unei opinii „sunt“ irelevante în raport cu rolul ei explicativ.

Acest principiu sugerează că problema specifică a reprezentării este greșit pusă, în sensul că, în cele din urmă, nu există vreo legătură reală între aspectele reprezentationale și cele explicative ale opiniei. Poate că modalitatea în care identificăm opiniile în termenii clauzelor de conținut (*content clauses*) având forma „că p“ creează aparența că rolul explicativ al opiniilor depinde, cumva, de capacitățile lor reprezentationale. Principiul (A) sugerează însă că asta este doar o iluzie, iar clauzele de conținut îndeplinesc două sarcini complet independente: pe de-o parte, ele precizează când anume opiniile sunt adevărate și când sunt false, iar pe de alta, indică semnificația explicativă a opiniilor.

Un răspuns uzual la acest argument constă în obiecția că, deși principiul (A) ar putea să evidențieze faptul că *valorile* de adevăr sunt irelevante în explicațiile acțiunilor, el nu dovedește totuși că și caracteristicile reprezentationale prezente sub forma condițiilor de adevăr ar fi irelevante în același fel. Dar acesta nu este un răspuns potrivit. În mod cert, este adevărat că potențialul explicativ al unei opinii depinde de conținutul ei – faptul de a fi, cum se spune, opinia că p sau opinia că q. Iar în acest sens, potențialul său explicativ depinde cu certitudine de „condițiile ei de adevăr“. Trebuie totuși să ni se spună ce are a face *adevărul* cu faptul de a avea o „condiție de adevăr“ în *sensul menționat*. Expresia „condiție de adevăr“ poate fi utilizată pentru a descrie cum să identificăm opiniile, pentru scopurile noastre explicative. Dar dacă principiul (A) este corect, identificările de acest tip ar servi la fel de bine scopurile explicative, chiar fără să specifice când ar fi opiniile adevărate și când ar fi false – cu alte cuvinte, chiar dacă ele nu ar fi câtuși de puțin identificări reprezentationale.

Astfel, principiul (A) pune la îndoială propunerea de a soluționa problema generală a reprezentării mentale pornind de la întrebarea specifică despre relația dintre reprezentarea mentală și explicația acțiunii, întrucât (A) sugerează că nu există nici o astfel de relație cu care să putem începe. Nu cred, totuși, că ar trebui să abandonăm în mod precipitat această presupunere și

să începem să căutăm în locul ei vreo explicație alternativă a reprezentării mentale. În ciuda caracterului persuasiv al argumentului cuprins în teza (A) – și el a convins mulți filosofi, inclusiv pe mine însumi<sup>4</sup> – concluzia lui trebuie privită, în mod sigur, cu suspiciune: este cert că nu poate fi o simplă *coincidență* faptul că aceeași manieră de identificare a opiniilor în termenii clauzelor „că p” ar fi urmată atât pentru scopurile reprezentationale, cât și pentru cele explicative. Dacă teza (A) implică această concluzie, atunci probabil că trebuie s-o reexaminăm.

În secțiunea următoare voi arăta că, contrar tezei (A), există o specie de explicație a acțiunii pentru care diferența dintre adevăr și falsitate contează. Concentrându-ne asupra acestei specii de explicație a acțiunii vom fi în stare să soluționăm problema specifică a felului în care reprezentarea mentală stă în relație cu explicația acțiunii. Iar asta ne va conduce înapoi către un răspuns la problema mai generală a felului cum este posibilă în genere reprezentarea mentală.

Înainte de a continua, aș vrea totuși să fac un scurt comentariu asupra unei teme adiacente. Mare parte din dezbaterea filosofică recentă s-a concentrat asupra opiniilor cu „conținuturi generale” (*broad contents*), opinii a căror corectă atribuire celor ce le susțin răspunde nu numai la ceea ce se petrece în capetele opinenților, dar și la ceea ce se petrece împrejurul acestora. Astfel, Hilary Putnam a susținut că identitatea opiniilor depre genurile naturale depinde de genurile care sunt prezente efectiv în lumea opinentului (1974); Tylor Burge era de părere că însuși conținutul opiniilor teoretice depinde de caracteristicile contextului social (1979, 1982), iar Gareth Evans a susținut că posesia opiniilor singulare reclamă existența obiectelor la care se referă acele opinii (1982). Toate aceste scenarii au ca deznodământ ideea că doi opinenți pot da fie identici moleculă-cu-moleculă și totuși, datorită diferențelor existente în mediile lor, unul poate să creadă că p, iar celălalt nu.

Unii filosofi par convinși că o dată admise opiniile generale, problemele filosofice ale reprezentării dispar. (Cf, Mc. Dowell, 1986, în special Secțiunea 5). Se sugerează aici că problemele reprezentării apar doar atât timp cât considerăm opiniile ca lucruri aflate în capetele oamenilor. Dar, o dată ce înțelegem că însăși posesia unei opinii poate să implice fapte extra-craniale, nu trebuie să ne mai îngrijoreze felul în care niște lucruri aflate în cap pot să trimită la lucruri din afara lui.

Acest mod de a gândi nu vizează cu adevărat problemele pe care le-am ridicat în secțiunea de față. Recunoscând că unele opinii au conținuturi *generale*, nu putem totuși să dăm seamă de faptul că ele posedă conținuturi *reprezentationale* ca atare. Haideți să acceptăm că posesia opiniilor poate să reclame, în unele cazuri, existența anumitor entități situate în afara opinenților. În orice caz, așa se întâmplă cu posesia unor stări precum solvabilitatea financiară, popularitatea sau faptul de a fi căsătorit. Solvabilitatea mea financiară nu reprezintă, prin urmare, lumea ca fiind într-un anumit fel. În consecință, de ce opiniile ar reprezenta lumea ca fiind într-un anume fel? Și ce are asta a face cu rolul lor explicativ? Aceste întrebări se mențin la fel de stringent în legătură cu opiniile generale, ca și cu oricare altele.

O chestiune similară îi vizează pe acei filosofi care pledează *împotriva* importanței psihologice a conținuturilor generale, pe temeiul că este dificil de înțeles cum ar fi posibil ca niște fapte din afara capului să aibă relevanță pentru semnificația explicativă a opiniilor (Cf. Fodor, 1987). Dacă aceasta este o îngrijorare legitimă cu privire la conținuturile generale, atunci există cu siguranță și o îngrijorare anterioară mult mai importantă – aceea de a înțelege cum anume conținuturile reprezentationale de orice fel pot să conteze în semnificația explicativă a opiniilor. Căci, lăsând de o parte cazurile speciale, diferența între opiniile adevărate și cele false este la fel de mult în afara capului ca și entitățile care fac generalitatea opiniilor generale.

Cred că, din întâmplare, o apreciere adecvată a conținuturilor generale emerge în mod natural dintr-o înțelegere satisfăcătoare a reprezentării mentale. În particular, odată ce înțelegem în ce fel conținutul reprezentational *per se* poate conta în explicația acțiunii, vom înțelege că nu e deloc surprinzător ca unele conținuturi să fie generale. Voi reveni la acest subiect la sfârșitul Secțiunii 4.

### 3. SEMNIFICAȚIA ADEVĂRULUI

De ce, inițial, este plauzibil să se accepte (A), principiul că valorile de adevăr sunt irelevante pentru rolul explicativ? Cred că

plauzibilitatea lui (A) derivă dintr-o anumită imagine asupra structurii explicației acțiunii:

- (B) 1. X dorește să t  
2. X crede că s va produce t  
De aceea  
3. X face s.

În mod evident, (B) ne furnizează numai o parte din ceea ce se solicită unor asemenea explicații mijloace-scop ale acțiunii. Există o sumedenie de cazuri când X face ceva, deși acest lucru nu poate fi explicat adecvat prin opinia lui X că acțiunea sa constituie un mijloc efectiv pentru realizarea scopului dorit. Spre exemplu, dorința cu pricina ar putea fi foarte slabă, sau X însuși ar putea să creadă că acțiunea s ar împiedica atingerea unui scop mai semnificativ. Din fericire, nimic din acest articol nu-mi cere să fiu mai concret în legătură cu ceea ce ar trebui adăugat la (B). Unii filosofi ar fi gata să susțină că dispunem de o explicație adecvată a acțiunii numai dacă opiniile și dorințele în chestiune sunt legate de acțiune prin vreo lege naturală. Alți filosofi ar spune că opiniile și dorințele trebuie să arate doar că acțiunea este rațională, într-un oarecare sens non-natural al termenului „rațional”. Pentru intențiile noastre prezente putem, totuși, să ocolim această temă și să considerăm, pur și simplu, că (B) reprezintă nucleul general acceptat al explicației acțiunii.

Este îndeajuns de evident de ce valorile de adevăr nu contează în explicațiile precum (B). Ceea ce explică (B) este *mijlocul*, s, adoptat de X. Dar opinia lui X este despre *eficacitatea* lui s. Dacă opinia este adevărată, atunci lucrurile se vor petrece așa cum se așteaptă X, iar dacă ea este falsă, nu. Însă în oricare din cazuri X face s. Așadar s este explicat la fel de bine indiferent dacă opinia din spatele acțiunii este adevărată sau falsă.

Putem astfel să atribuim plauzibilitatea principiului (A) structurii explicative (B). De aceea principiul (A) nu este adevărat fără rezerve. Uneori dorim să explicăm faptul că agenții își ating scopurile, mai degrabă decât să adoptă anumite mijloace. Iar atunci valorile de adevăr devin relevante.

Să presupunem că mi se strică televizorul. Îi telefonez fratelui meu, iar el vine și îl repară. Dumneavoastră sunteți nedumeriți în legătură cu capacitatea lui de a face asta. Eu vă spun că a citit o carte bună pe această temă, și de aceea posedă informații despre care anume reglaje vor remedia anume defecte. În felul acesta eu vă explic succesul lui în repararea televizorului.

Explicându-i succesul, eu invoc opinia lui că un anumit reglaj va duce la repararea televizorului. Dar acum, pentru că explic atingerea scopului său și nu adoptarea unui mijloc, adevărul acestei opinii nu este relevant. Dacă opinia pe care a căpătat-o din manual nu era *adevărată*, atunci faptul că avea acea opinie nu ar putea explica faptul remedierii defectiunii, ci numai efectuarea reglajelor în chestiune.

Făcând apel la adevărul opiniei fratelui meu, pot să explic atingerea scopului dorit de el, și nu doar faptul că a adoptat un anumit mijloc. Altfel spus, explicația mea are următoarea structură (C), mai degrabă decât structura (B):

- (C) 1. **X** dorește să **t**
- 2. **X** crede că o anumită acțiune va produce **t**
- 3. Această opinie este adevărată
- De aceea
- 4. **X** produce **t**.<sup>5</sup>

#### 4. ADEVĂRUL CA GARANȚIE A SUCCESULUI

Explicațiile precum (C) arată de ce diferența dintre opiniile false și cele adevărate contează în explicația acțiunii. Dacă explicațiile acțiunilor s-ar ocupa numai de mijloacele agentului, precum în (B), atunci forța reprezentatională ar fi irelevantă. Dar, o dată ce suntem preocupați de explicarea reușitei, precum în (C), semnificația explicativă a unei opinii depinde tocmai de adevărul sau falsitatea ei.

Obținem astfel o soluție la problema specifică a reprezentărilor mentale, problema relației dintre forța reprezentatională a



opiniilor și rolul lor explicativ: schema explicativă (C) arată că a cunoaște condiția de adevăr a unei opinii înseamnă tocmai a ști când poate ea să explice succesul. Dar, poate oare să ne ducă acest fapt mai departe, spre o soluție la problema generală a înțelegerii felului în care este posibilă reprezentarea mentală într-o lume intrinsec non-reprezentatională? Mi-ar plăcea să sugerez că poate. Adică, mi-ar plăcea să sugerez că prin înțelegerea felului în care reprezentarea intră în legătură cu explicația acțiunii, putem să înțelegem *ce este* reprezentarea: și tocmai acest aspect al opiniilor este cel care le permite să explice succesul acțiunilor.

Permiteți-mi să formulez această sugestie ca o analiză explicită:

- (D) Condiția de adevăr a oricărei opinii este acea condiție care garantează că acțiunile bazate pe opinia respectivă vor reuși. (O opinie este, așadar, adevărată dacă este realizată condiția ei de adevăr.)

Soluțiile problemelor filosofice nu trebuie să ia în mod necesar forma unei analize explicite a conceptelor. Nu există însă nimic în neregulă cu analizele explicite, dacă le puteți formula. Așa cum am spus în Introducere, din întâmplare eu nu cred că analiza condițiilor de adevăr în termenii garanției succesului (*success-guaranteeing analysis of truth conditions*) este completă până când nu este plasată într-un context teleologic. Dar cred, totuși, că ea reprezintă *o parte* a adevărului despre condițiile de adevăr. Așa că, în acest stadiu, ar fi util să punem în discuție unele din obiecțiile inițiale la (D) și să facem câteva observații preliminare. Nevoia de teleologie va fi argumentată în secțiunea următoare.

(1) Oare (D) nu se aplică doar la opiniile de forma: *s* va produce *t*? Într-adevăr, acestea sunt singurele opinii ce au fost, până acum, prezentate ca relevante pentru succesul acțiunilor. Însă, desigur, o analiză a condițiilor de adevăr trebuie să se ocupe de toate formele de opinii, și nu numai de cele referitoare la mijloacele adaptate scopurilor.

Nu este, totuși, dificil de înțeles de ce (D) ar trebui să fie considerat ca referindu-se la opiniile de orice formă, la fel ca și la opiniile despre mijloace-scopuri. Este adevărat că, în ultimă instanță, relevanța opiniilor pentru acțiuni depinde întotdeauna

de ceea ce implică ele în legătură cu mijloacele adecvate. Iar în acest sens, numai opiniile despre mijloace-scopuri sunt relevante în *mod direct* pentru acțiuni. Dar, mai mult decât atât, asemenea opinii despre mijloace-scopuri (s va produce t) vor fi, de regulă, inferate de către agent din diverse alte opinii. Însă atunci, tocmai acestea vor institui legătura generală care se solicită să existe între adevăr și reușită. Căci, dacă acele alte opinii sunt adevărate, iar inferențele bazate pe ele sunt valide, atunci opinia că s va produce t va fi de asemenea adevărată, iar acțiunea rezultată va reuși. Astfel, faptul că acțiunile bazate pe opinii adevărate vor avea succes reprezintă un principiu *general*, și nu doar un principiu relativ la opiniile despre mijloace-scopuri ca atare.<sup>6</sup> În consecință, atunci când răsturnăm acest principiu într-o analiză a condițiilor de adevăr – analiza (D) –, aceasta promite să se aplice opiniilor în general, și nu doar opiniilor de forma mijloace-scop.

(2) În general, în spatele oricărei acțiuni date se vor afla *mai multe* opinii. Dar asta înseamnă că adevărul unei singure opinii va fi insuficient pentru garantarea succesului acțiunilor rezultate. Căci succesul va fi garantat numai dacă celelalte opinii din spatele acțiunii sunt de asemenea adevărate. Așadar, analiza (D) va trebui reformulată în mod strict astfel:

Condiția de adevăr a oricărei opinii este acea condiție care garantează că acțiunile bazate pe respectiva opinie vor reuși, presupunând că oricare *alte* opinii în concordanță cu care se acționează vor fi de asemenea adevărate.

Dar atunci, această caracterizare descalifică (D) ca analiză a reprezentării condiționate-de-adevăr (*truth-conditional representation*), întrucât ea presupune noțiunea de adevăr în explicarea acesteia.

S-ar părea că am putea să tratăm această dificultate considerând că analiza (D) s-ar aplica în mod specific cazurilor în care opiniile singulare generează acțiuni pe cont propriu, fără asistența altor opinii<sup>7</sup>. Condițiile de adevăr ar putea fi atunci identificate cu ceea ce garantează succesul în acele cazuri când avem de-a face cu astfel de opinii singulare. Necazul cu această abordare este că ajungem iarăși la obiecția (1), din moment ce singurul fel de opinii ce pot genera acțiuni proprii sunt opiniile mijloace-scopuri.<sup>8</sup> Dacă dorim o analiză a adevărului care să

funcționeze pentru opinii în general, și nu doar pentru opiniile mijloace-scopuri, atunci ne trebuie o cale de a extinde (D) dincolo de cazul alegerii acțiunii, luând ca bază o singură opinie.

O cale mai bună de a trata problema în cauză este să considerăm că analiza (D) s-ar aplica simultan la toate tipurile de opinii aflate în repertoriul unui agent. Altfel spus, ar trebui să considerăm că (D) fixează condițiile de adevăr pentru toate acele opinii privite colectiv, prin ceea ce am putea numi „rezolvarea” unui set de ecuații simultane. „Ecuatiile” sunt presupunerile conform cărora condiția de adevăr a fiecărei opinii va garanta succesul acțiunii, dacă alte opinii relevante vor fi adevărate; iar atunci, „soluția” generală este o asumare colectivă a condițiilor de adevăr ce satisfac toate aceste ecuații.

(3) O altă îngrijorare inițială cu privire la (D) ar putea fi aceea că ea face adevărul prea accesibil. Desigur, noi nu vrem să socotim opiniile ca adevărate ori de câte ori acțiunile pe care le sugerează au rezultate satisfăcătoare. Poate oare o acțiune să conducă, numai printr-un simplu accident, la un rezultat dorit, adică, chiar dacă unele opinii din spatele ei sunt false (ca și cum ele ar implica o anume greșeală ce se auto-corectează)?

Însă (D) nu exclude, de fapt, această posibilitate. Nu ni se sugerează însă că pentru a putea vorbi de adevărul unui întreg set de opinii exprimate ar fi suficient să reușească o singură acțiune particulară, sugerată de acele expresii particulare. Mai degrabă, (D) specifică o condiție care *garantează totalității* expresiilor de tipurile relevante că acțiunile rezultate din ele vor avea succes.

(4) Ce se poate spune despre deciziile luate în condiții de incertitudine? În multe situații, un agent va acționa bazându-se nu pe opinii complete, ci pe unele parțiale. În astfel de cazuri gândirea agentului nu va alege vreo acțiune pentru că ea ar avea șanse de succes certe, ci, mai degrabă va selecta acțiunea care va avea cea mai mare probabilitate subiectivă de succes. Însă atunci, chiar dacă acțiunea reușește *realmente*, ea nu a fost garantată de faptul că opiniile agentului ar fi reprezentat corect lumea.

O întrebare interesantă se referă la cât de mult depinde buna fundamentare a deciziilor în condiții de incertitudine de caracteristicile obiective ale lumii, cum ar fi existența șanselor obiective. Putem însă să ocolim această temă aici. Căci, încă o dată, nu

există în (D) nimic care să excludă posibilitatea acțiunilor al căror succes *nu este* garantat de adevărul opiniilor aflate în spatele lor. Ideea din subtextul lui (D) este mai degrabă aceea că ne-am putea concentra asupra acelor cazuri în care succesul *este* garantat astfel, urmând apoi să analizăm adevărul ca fiind ceea ce garantează succesul tocmai în acele cazuri. Prin urmare deciziile în condiții de incertitudine rezultate din opinii parțiale sunt în afara discuției. Pentru a aplica (D) la o opinie dată ar trebui să rămânem la cazurile în care acea opinie este susținută *complet*, și la procedurile decizionale care *nu sunt* incerte: în *acele* cazuri, succesul este garantat de adevăr<sup>9</sup>.

(5) Analiza (D) pare să implice că virtutea adevărului este esențialmente pragmatică și că motivul pentru care dorim adevărul este acela că urmărim întotdeauna succesul acțiunilor noastre. Dar cu siguranță adevărul poate fi urmărit și ca scop în sine, și nu doar din pricina valorii sale pragmatice. Într-adevăr, există anumite probleme – referitoare, să zicem, la înțelegerea mai detaliată a universului, sau la trecutul îndepărtat – în care interesul nostru în posedarea unor opinii adevărate nu poate fi unul practic, de vreme ce asemenea opinii nu pot avea importanță pentru acțiunile noastre.

Însă o asemenea întâmpinare își ratează ținta. (D) nu este o teorie depre motivul pentru care ar trebui să *dorim* adevărul. Ea e o teorie despre *ce este* adevărul: anume, în raport cu o opinie, realizarea unei condiții care să garanteze că, dacă ar fi ca un agent să acționeze bazându-se pe acea opinie, acțiunea rezultată i-ar reuși<sup>10</sup>. Asta nu presupune, totuși, că toată lumea va acționa efectiv bazându-se pe acea opinie, și nu presupune nici că singurul motiv pentru a dori adevărul în legătură cu opinia respectivă ar fi acela că, bazându-se pe ea, ai fi în stare să acționezi cu succes. Fără doar și poate, dacă într-adevăr vrei să reușești într-o acțiune, atunci (D) implică imediat că ai un motiv pentru a vrea ca opiniile din spatele ei să fie adevărate. Dar asta lasă loc altor motive de a vrea să deții adevărul, atât în cazul opiniilor semnificative sub aspect practic, cât și în al celor practic-nesemnificative. În particular, lasă loc valorizării adevărului ca scop în sine. (N-am putea oare să ne întrebăm acum: *de ce* să fie adevărul valorizat ca un scop în sine? Dar iau ca pe o virtute a lui (D) faptul că ne permite să considerăm că această întrebare ar avea sens).

(6) Analiza (D) vorbește despre „condiții de adevăr” și de faptul că asemenea condiții ar putea fi „realizate”. Dar nu ne angajăm, atunci, în afirmarea unor entități dubioase, precum propozițiile, stările de lucruri posibile sau mulțimile de lumi posibile?

Unii filosofi n-ar fi afectați de asemenea angajamente. Ei pot sări direct la următoare obiecție. Dar, pentru cei afectați, dați-mi voie să încerc să arăt că reificarea condițiilor de adevăr nu este esențială pentru (D).

Să considerăm următoarea schemă:

(E) O acțiune bazată pe opinia<sub>n</sub> va avea succesul garantat doar în cazul *p*.<sup>11</sup>

Putem înțelege analiza (D) în sensul că ar aserta faptul că afirmațiile de forma (E) specifică conținuturile condiționate-de-adevăr ale opiniilor. Asta înseamnă că analiza (D) poate fi înțeleasă ca asertând faptul că (E) este un substitut echivalent pentru.

(F) Opinia<sub>n</sub> este adevărată dacă și numai dacă *p*.

Observați acum că nici (E) și nici (F) nu se referă la condițiile de adevăr ca atare. Deci, dacă semnificața analizei (D) este pur și simplu aceea că (E) este echivalent cu (F), atunci analiza (D) va fi eliberată, de asemenea, de orice angajament substanțial față de condițiile de adevăr.<sup>12</sup> (Voi adopta de aici înainte următoarea accepțiune a analizei (D): referințele subsecvente la condițiile de adevăr trebuie privite ca niște conveniențe expositive.)

Ce putem spune acum despre adevărul însuși? Ei bine, pentru orice opinie particulară, instanța relevantă a lui (F) ne va spune direct ce anume se cere pentru ca acea opinie să fie adevărată: de exemplu, considerând opinia că *zăpada este albă*, ea va fi adevărată numai în cazul că *zăpada este albă*; dacă vom considera opinia că *iarba este verde*, atunci ea va fi adevărată numai în cazul în care *iarba este verde* – și așa mai departe. Dar, lucru și mai interesant, cum analiza (D) ne dă o evaluare a semnificației cerințelor de forma (F) – o evaluare ce nu presupune, prin ea însăși, noțiunea de adevăr – putem generaliza, spunând că *orice* opinie va fi adevărată numai în cazul în care acțiunile bazate pe acea opinie vor avea succesul garantat.<sup>13</sup>

Aceste ultime remarci ne ajută să înțeleagă *ce tip* de analiză ni se oferă prin (D). Evident, (D) nu este o analiză *conceptuală*: conceptul pre-teoretic de adevăr nu este același lucru cu conceptul garantării succesului. Mai degrabă (D) ne oferă un fel de analiză *reductivă* a adevărului: ea *ne explică* ce este în realitate adevărul, chiar dacă noi nu ne gândim la el în acest mod. Dar care este atunci baza reductivă? În mod cert, (D) nu pune semnul egalității între adevăr și vreo proprietate fizică (sau vreo altă proprietate de ordinul întâi) naturală. Aș sugera că noi privim (D) ca și cum ar identifica adevărul cu un fel de proprietate naturală de *ordinul doi*: proprietatea opiniei de a *cauza* acțiuni garantate că *ar cauza* rezultate reușite. Avem de-a face aici cu un analog al definițiilor funcționaliste, în termenii unor roluri cauzale, pentru stări de ordinul doi. La fel cum o definiție funcționalistă definește o stare invocând o teorie care specifică rolul cauzal al acelei stări, tot așa eu sunt de părere că ar trebui să definim adevărul prin referire la propria noastră teorie a semnificației cauzale a adevărului, adică, prin referire la teoria că acțiunile cauzate de opinii adevărate cauzează succesul. Analogia poate fi dusă mai departe: așa cum o stare definită funcțional poate fi „realizată” de stări fizice de ordinul întâi în diferite sisteme, tot așa adevărul va fi „realizat” în mod diferit pentru opinii diferite: în cazul opiniei că zăpada este albă, adevărul (adică proprietatea cauzării acțiunilor care vor cauza rezultate reușite) este realizat prin faptul că zăpada este albă; în cazul opiniei că iarba este verde, el este realizat prin faptul că iarba este verde – și așa mai departe.

Între abordarea pe care o apăr eu aici și „abordarea redundantă” a adevărului există anumite similitudini, constând în aceea că eu susțin că în cazul unei opinii nu avem nevoie să ne gândim la adevăr în termenii „realizării condiției de adevăr” a opiniei, ci pur și simplu în termenii existenței stărilor de lucruri pretinse de opinii. Există însă similitudini și cu „abordarea prin corespondență”, prin faptul că, atunci când explic ce este conținutul reprezentational al unei opinii – adică, atunci când explic ce înseamnă ca o opinie să reclame ceea ce reclamă –, asta implică specificarea felului în care trebuie să fie *lumea* pentru ca acțiunile bazate pe acea opinie să reușească. Chiar dacă în acest fel nu sunt furnizate condițiile de adevăr ca atare, adevărul este făcut să depindă de felul cum este lumea în mod independent.<sup>14</sup>

(7) Analiza (D) se aplică numai la opiniile al căror adevăr dispune de o potențială relevanță causală pentru succesul acțiunilor. Poate că asta îi permite să împacă opiniile despre lumea naturală. Dar cum rămâne cu judecățile morale, modale ori matematice? În ce sens – dacă e posibil vreunul – poate să conteze pentru reușita acțiunii adevărul unor asemenea judecăți non-naturale?

Nu-mi propun să urmăresc aici această temă complexă. Faptul că analiza (D) s-ar putea sau nu aplica la o anumită categorie de judecăți depinde de detaliile funcționării acestora, iar asemenea detalii constituie subiectul unei aprinse controverse atunci când e vorba de judecățile morale, modale sau matematice. Părerea mea de moment este că (D) poate fi făcut să se aplice la acele domenii ale discursului care sunt expresii genuine ale *opinieii*, dar că este discutabil cât de bine se potrivesc în această matrice discursurile morale, modale sau matematice. Însă, așa cum am spus, asta e o temă prea complexă pentru a fi urmărită aici.

(8) În răspunsul meu la obiecția (1) am apelat la noțiunea de *validitate*: am susținut că analiza (D) s-ar putea extinde de la opiniile mijloace-scopuri la alte opinii, deoarece inferențele valide bazate pe orice fel de opinii adevărate vor conduce la concluzii adevărate despre mijloacele adecvate. S-ar putea susține, totuși, că acest apel la validitate este ilegitim, pe temeiul că noțiunea de validitate presupune noțiunea de adevăr.

Analiza (D) are în mod cert *nevoie* de noțiunea de validitate. Agenții vor produce adesea inferențe *nevalide* cu privire la mijloace (imaginați-vă că trebuie să decidă imediat ce anume să facă, sau că situația lor este foarte complicată), iar atunci adevărul opiniilor pe care se bazează acele inferențe *nu va garanta* reușita acțiunilor lor. Așadar, dacă este ca (D) să se aplice în general, și nu doar la opinii mijloace-scop, în mod strict el va trebui formulat astfel:

Condiția de adevăr a oricărei opinii este condiția care garantează faptul că acțiunile bazate în mod valid pe acea opinie vor reuși.

În acest fel, problema se clarifică: dacă (D) presupune validitatea, iar validitatea presupune adevărul, cu greu s-ar mai putea susține că ea ar constitui o analiză a adevărului.

O posibilă mișcare ar consta, acum, în negarea faptului că validitatea depinde realmente de adevăr. Am putea astfel să căutăm o noțiune pur sintactică a validității (definită în termenii unei structuri de reguli inferențiale), mai degrabă decât noțiunea semantică a unei forme inferențiale care să conducă în mod garantat de la adevăr la adevăruri. Această strategie sintactică pare totuși nepromițătoare. Mai întâi, există dificultăți tehnice legate de completitudinea caracterizărilor sintactice ale validității care-nu-este-de-ordinul-întâi. Apoi, în orice caz, dat fiind că astfel de caracterizări sintactice corespund întotdeauna concepției semantice asupra validității<sup>15</sup> – chiar și pentru validitatea de ordinul întâi –, e îndoielnic că strategia sintactică ar elimina realmente circularitatea, mai degrabă decât s-o ascundă sub covor.

Din aceste motive sunt de acord că noțiunea de validitate ridică o dificultate substanțială în fața analizei (D). Cred că această dificultate poate fi surmontată, dar nu câtă vreme nu vom fi plasat (D) într-un context teleologic. Așa că haideți să o lăsăm deoparte până la secțiunea următoare.

Observațiile de până aici completează catalogul obiecțiilor mele inițiale la adresa analizei (D). Însă înainte de a continua, permiteți-mi să completez această secțiune revenind, pe scurt, la tema conținuturilor generale.

Dacă în gândirea noastră opiniile au ca unic rol să explice adoptarea mijloacelor, ca în cazul

- (B) 1. X dorește să t  
2. X crede că s va produce t  
De aceea  
3. X face s.

atunci n-ar exista nici un motiv pentru care vreo opinie să fie generală. Căci explicațiile de forma (B) nu reclamă ca opiniile să facă altceva decât, cum s-ar spune, să dea acțiunilor un *imbold* causal din interior. Iar potrivit acestei concepții asupra opiniei, ar fi într-adevăr stânenitor ca diferențele din afara capetelor opinenților să poată modifica ceea ce cred ei.

Din perspectiva analizei (D) putem, totuși, să înțelegem că această concepție asupra opiniei lasă deoparte ideea esențială a conținutului de adevăr condițional. Conținutul de adevăr condi-



țional nu are nimic de-a face cu „îmboldurile“ interne<sup>16</sup>. Ele specifică, mai degrabă, condițiile de succes pentru ceea ce se întreprinde: deci o condiție care ne asigură că mijloacele adoptate vor produce scopul dorit. O dată ce ne este clar acest lucru, e ușor de înțeles de ce ar trebui ca unele opinii să aibă conținuturi dependente de lume. Aici avem de-a face pur și simplu cu ideea că, dacă vorbim despre doi oameni identici din punct de vedere fizic, aflați însă în lumi diferite, atunci ceea ce este necesar pentru ca acțiunea uneia dintre persoane să reușească poate să difere de ceea ce se cere pentru succesul acțiunii celeilalte persoane.

Acest aspect este exprimat cel mai clar în cazul opiniilor indexiale explicite. Să presupunem că Bill și Ben sunt fizic identici, și că ambii au dorința și opinia pe care o exprimă prin „Vreau să-mi fie cald“ și prin „Alergarea pe loc mă face să mă încălzesc“. Probabil că în acest caz amândoi vor începe să facă aceleași mișcări ale corpului – și anume, să alerge pe loc. Sub acest aspect opiniile lor sunt similare: ambele opinii „îmboldesc“ în același mod. Remarcați însă că aici condițiile ce vor satisface dorințele lor sunt *diferite*: dorința lui Bill va fi satisfăcută de încălzirea *lui Bill*, pe când dorința lui Ben va fi satisfăcută de încălzirea *lui Ben*. De aceea vor fi diferite și condițiile cerute pentru ca acțiunile lui Bill și Ben să reușească: acțiunea lui Bill va reuși doar dacă alergarea *lui Bill* îl va încălzi *pe Bill*, în vreme ce reușita acțiunii lui Ben reclamă condiția – complet diferită – ca alergarea *lui Ben* să-l încălzească *pe Ben*. În conformitate cu (D), acesta e motivul pentru care condițiile de adevăr ale opiniei lui Bill și Ben sunt diferite, în ciuda identității fizice a celor doi. Asta se datorează, pur și simplu, faptului că acțiunile lui Bill și Ben au condiții de succes diferite.

Nu e greu de înțeles cum s-ar putea extinde această poveste pentru a fi aplicată la opinii care să nu fie explicit indexiale, așa cum sunt opiniile despre genurile naturale. Dacă presupunem că acea condiție ce satisface dorința exprimată prin „apă“ depinde de tipul de lume în care se află dorința, atunci condiția cerută pentru succesul acțiunilor ce ținteau acea dorință va depinde, în mod similar, de lumea agentului. Și astfel, în conformitate cu analiza (D), se va putea explica de ce opiniile despre „apă“ au condiții de adevăr dependente de lume.

După cum se vede, scenariul din urmă acceptă ca dat faptul că *dorințele* care implică concepte ale genurilor naturale au condiții generale de *satisfacere*. Dar poate că acest lucru e discutabil. În ce privește dorințele explicit indexiale, cum ar fi aceea exprimată prin „Vreau să-mi fie cald“, este destul de normal ca posibila condiție de satisfacere a dorinței să depindă de cel ce o resimte. Nu mai este însă la fel de clar de ce ar trebui să depindă condiția de satisfacere a dorinței exprimate prin „apă“ de tipul substanței chimice prezente în lumea celui ce o resimte, mai degrabă decât de existența transmundană a unui lichid incolor, inodor și insipid. Permiteți-mi însă să abandonez tema respectivă în acest stadiu. Am arătat cel puțin că principiul (D) oferă o explicație a opiniilor generale, *dacă* presupunem că există dorințe generale. În finalul secțiunii următoare mă voi întoarce la dorințele generale însele.

## 5. TEORIA TELEOLOGICĂ A REPREZENTĂRII

Tocmai am văzut că întrebările ce vizează condițiile de satisfacere a dorințelor pun în dificultate explicarea conținuturilor generale, așa cum este ea sugerată de analiza (D). Vreau să arăt acum în ce fel tema condițiilor de satisfacere pune, de asemenea, într-o dificultate mai generală analiza (D) în întregul ei.

Să ne reamintim că, în conformitate cu (D), adevărul este cel ce garantează *succesul*. Până acum n-am spus prea multe despre succes ca atare, dar din ceea ce am spus rezultă cu claritate că nu e vorba de o noțiune primitivă. Ea derivă, mai degrabă, din noțiunea de *dorință* ce este *satisfăcută*: în sensul vizat, o acțiune reușește doar atunci când este îndeplinită condiția de satisfacere a dorinței ce stă în spatele acțiunii.

În ce privește dorințele, condițiile de satisfacere și satisfacerea prezintă o strânsă analogie cu ceea ce reprezintă pentru opinii condițiile de adevăr și adevărul. La fel cum fiecare opinie are o condiție de adevăr care, împreună cu celelalte trăsături ale lumii, determină o valoare de adevăr a opiniei, tot așa fiecare

dorință are o condiție de satisfacere care, alături de celelalte trăsături ale lumii, determină dacă dorința este sau nu satisfăcută.

Am început acest articol susținând că, din punct de vedere filosofic, faptul că opiniile ar avea condiții de adevăr în virtutea cărora ele ar reprezenta lumea ca fiind într-un anume fel constituie o enigmă. Aș fi putut totuși să susțin, la fel de bine, că sub aspect filosofic la fel de enigmatic ar fi și ca dorințele să aibă condiții de satisfacere în virtutea cărora ele să „ceară“, ca să zicem așa, ca lumea să fie într-un anumit fel.

Dar atunci este cu siguranță nesatisfăcător să oferim problemei reprezentării – privitoare la opinii – o soluție care, așa cum se întâmplă în (D), ia de bune proprietățile reprezentative ale dorințelor. Dacă asta e tot ce putem oferi ca explicație filosofică, nedumerirea cu care am început va persista.

N-am putea oare, pur și simplu, să analizăm satisfacerea pe aceleași coordonate pe care am analizat adevărul? Pentru a vedea cum s-ar putea întâmpla asta, dați-mi voie să clarific, printr-un enunț explicit, legătura dintre adevăr și satisfacere:

(G) Acțiunile bazate pe opinii adevărate vor satisface dorințele spre care tind.

În fapt, noi am derivat (D) concentrându-ne asupra rolului adevărului în acest principiu și „răsturnând“ apoi principiul pentru a efectua o analiză a ceea ce *este* adevărul. Acum ni se sugerează că am putea să facem același lucru cu satisfacerea, adică am putea pur și simplu să „răsturnăm“ (G) încă o dată – dar în legătură cu dorințele, mai curând decât cu opiniile – pentru a obține:

(H) Condiția de satisfacere a unei dorințe este acea condiție despre care se garantează că rezultă din acțiunile bazate pe dorința respectivă, dacă opiniile din spatele acțiunii sunt adevărate.

Dar acest principiu nu funcționează. Încercarea de a face o analiză a reprezentării, atât pentru opinii cât și pentru dorințe, pornind de la (G), e asemenea încercării de a rezolva o singură ecuație cu două necunoscute. Problema e că va exista un număr

nedeterminat de modalități de a atașa condițiile de „adevăr” și „satisfacere” la opiniile și dorințele unui agent, toate răspunzând constrângerii ca „adevărul” să garanteze „satisfacerea”, dar numai una dintre ele fiind în stare să identifice conținuturile reprezentationale reale ale acelor opinii și dorințe. (Cf. Stalnaker, 1984, pp. 17–18; Papineau, 1984, p. 555; Papineau, 1987, p. 57). Așa că, pentru a explica ce anume determină corectitudinea identificării corecte, trebuie să adăugăm ceva la explicația reprezentării așa cum e oferită de principiul (G) luat separat.

În acest punct, o sugestie naturală ar fi să suplimentăm coordonarea mutuală a adevărului și satisfacerii, cerută de (G), cu ideea – care nu a apărut până acum în acest articol – că este *normal* ca opiniile să fie prezente atunci când sunt asigurate condițiile lor de adevăr, și este *normal* ca dorințele să-și genereze propriile condiții de satisfacere.

Ce-ar trebui să înțelegem, în acest caz, prin „normal”? O posibilitate ar fi să-i dăm o interpretare *statistică*. Condițiile de adevăr ale opiniilor ar apărea atunci ca fiind acele circumstanțe prezente *cel mai frecvent* atunci când sunt susținute opiniile respective. Iar în cazul dorințelor, condițiile de satisfacere ar apărea atunci ca fiind acele circumstanțe care se produc *cel mai frecvent* atunci când dorințele respective sunt urmate.

Se recunoaște totuși, în general, că acest tip de abordare statistică nu funcționează. (Vezi Fodor, 1984; Papineau, 1984; Dretske, 1986). Principala dificultate este că multe alte condiții vor fi co-prezente, alături de opinii și dorințe, la fel de frecvent ca și condițiile lor de adevăr și satisfacere. Să considerăm astfel condiția ce constă în *disjunctia* dintre condiția de adevăr a unei opinii date și circumstanțele care, în mod caracteristic, îi ademenesc pe oameni să-și mențină o opinie chiar și atunci când ea este falsă. Această condiție disjunctivă va fi co-prezentă alături de opinia respectivă chiar *mai frecvent* decât condiția de adevăr a opiniei, deoarece condiția respectivă acoperă nu numai toate cazurile în care opinia e adevărată, dar și unele dintre acelea în care ea e falsă. Ceea ce dovedește că, oricare ar fi condițiile de adevăr, ele nu sunt acele circumstanțe co-prezente cel mai frecvent alături de opinii. O constatare similară se aplică și în cazul dorințelor. Oricare ar fi dorința, nu este dificil să construiești condiții care să-i urmeze chiar mai frecvent decât este îndeplinită

ndiția ei de satisfacere, fapt ce respinge teza potrivit căreia condițiile de satisfacere a dorinței sunt acele condiții care o soțesc cel mai frecvent.

Acesta este punctul în care cred că avem nevoie de teleologie. Perspectiva de a defini adevărul și satisfacerea în termeni pur statistici este redusă. Dacă introducem însă noțiunea de *scop* (*purpose*) problema devine relativ simplă. Putem spune atunci că în condiția de adevăr a unei opinii avem în vedere acea condiție în raport cu care scopul opiniei este acela de a-i fi co-prezentă. Scopul unei opinii nu este acela de a fi prezentă atunci când se realizează anumite circumstanțe înșelătoare, și din acest motiv aceste circumstanțe nu pot fi parte a condițiilor ei de adevăr. Similar, *scopul* dorințelor nu este de a da naștere altor circumstanțe decât condițiilor lor de satisfacere.

Nu este aici locul potrivit pentru a oferi o apărare detaliată a acestei teorii teleologice a reprezentării, dar pot fi abordate două probleme de ordin general.

Mai întâi, este clar că valoarea acestei teorii depinde de o eventuală nouă evaluare pe care s-o putem da noțiunii teleologice de „scop“. Căci fără o asemenea evaluare, scopurile par la fel de dubioase din punct de vedere metafizic ca și capacitățile reprezentationale. Totuși, putem să facem scopurile metafizic acceptabile, socotindu-le în termenii proceselor de selecție. În această concepție, ar fi adecvat să spui că scopul lui A este să producă B în eventualitatea că prezența lui A s-ar datora unei operațiuni trecute, aparținând unui proces de selecție ce a determinat selectarea itemilor care să ducă la producerea lui B. (Cf. Wright, 1973 și Neander, 1984).

Dacă adoptăm această evaluare selecționistă a scopurilor, atunci cerințele referitoare la scopurile opiniilor și dorințelor nu vor fi, din punct de vedere metafizic, mai dubioase decât cerințele biologice precum: „Scopul inimii este să pompeze sângele“ sau „Scopul tremurăturii este să mărească temperatura corpului“. Ideea ar fi că opiniile au fost selectate deoarece sunt în mod caracteristic co-prezente alături de anumite condiții care, de aceea, trec drept condițiile lor de adevăr, și că dorințele au fost selectate deoarece în mod caracteristic ele dau naștere anumitor rezultate care, din acest motiv, trec drept condiții de satisfacere a lor.

Poate ar merita să subliniem că această abordare selecționist-teleologică a reprezentării mentale *nu* implică ideea că toate abilitățile reprezentationale trebuie să fie produse genetic înăscute ale selecției inter-generaționale. Căci teleologia bazată pe selecție poate să fie, de asemenea, produsul unei învățări individuale (cf. „Porumbelul apasă pe stinghie pentru a obține hrana”). Și în acest fel, dacă vreo opinie sau vreo dorință ne-înăscută este selectată în cursul învățării individuale, în virtutea co-prezenței condiției sale, sau în virtutea rezultatului pe care îl generează, atunci acea opinie sau dorință ar avea un veritabil scop reprezentational bazat pe selecție, în ciuda faptului că ele nu ar fi înăscute<sup>17</sup>.

A doua idee generală pe care aș vrea s-o semnez în legătură cu teoria teleologică a reprezentării este că teoria respectivă *nu înlocuiește* anterioara abordare a adevărului prin garanția succesului, ci mai degrabă o *încorporează*.

Pentru a înțelege asta ar fi util să ne gândim la opinii și dorințe ca la niște componente ale unui sistem decizional general uman și, în mod corespunzător, să considerăm că scopurile opiniilor și dorințelor derivă din contribuția lor la scopul general al acestui sistem.

Se poate considera că scopul general al sistemului decizional uman generează acțiuni care cauzează rezultatele corespunzătoare. Atât opiniile cât și dorințele contribuie la acest scop. Totuși, ele contribuie *în modalități diferite*. Rolul dorințelor este de a gestiona faptul că *rezultate* diferite corespund unor momente diferite: dorințele noastre variază pentru ca acțiunile noastre să producă rezultate diferite în momente diferite. Rolul opiniilor este să se ocupe de faptul că, dat fiind un rezultat oarecare, *în momente diferite* sunt adecvate aceleiași rezultat *mijloace* diferite: opiniile noastre variază pentru că putem alege, în diferite momente, mijloacele cele mai eficiente în raport cu rezultatele pe care le dorim într-un anumit moment.

În fine, toate scopurile bazate pe selecție depind de *rezultate*: a avea un scop înseamnă a-l fi selectat printr-un mecanism ce favorizează anumite *rezultate*. Aceste ultime remarci dovedesc, totuși, că opiniile se potrivesc schemei în cauză numai într-un fel cam complicat. Căci opiniile nu au rezultate proprii. Scopurile lor

sunt mai degrabă acelea de a produce orice fel de rezultate care ar satisface dorințele alături de care acționează. În realitate, opiniile selectate la un moment dat se modifică, în virtutea faptului că se dovedesc potrivite să cauzeze acțiuni care, la rândul lor, cauzează rezultatele dorite.

Asta implică faptul că trebuie să considerăm că abordarea teleologică a reprezentării se aplică mai întâi la dorințe și la satisfacere. Orice dorință dată va fi prezentă pentru a produce un anumit rezultat  $r$ , care reprezintă, din acest motiv, condiția de satisfacere a dorinței. Acceptând această explicație a satisfacerii și a dorinței, putem să explicăm și scopurile opiniilor. Orice opinie dată va fi prezentă pentru a produce acțiuni care, la rândul lor, vor produce rezultatele dorite, în cazul în care va exista condiția  $p$ , ce va reprezenta astfel condiția de adevăr a opiniei.

Acesta este doar un alt fel de a spune că în raport cu o opinie, condiția de adevăr este acea condiție care garantează că acțiunile pe care le va produce vor satisface dorințele. Dar asta nu e altceva decât analiza (D) menționată anterior, adică analiza condițiilor de adevăr prin garanția succesului. Teleologia nu a adăugat decât o evaluare independentă, în termenii scopurilor biologice, a satisfacerii dorinței. Dată fiind această evaluare a satisfacerii dorinței, opiniile pot fi înțelese ca având la fel de bine și scopuri biologice – anume, acelea de a permite dorințelor să-și îndeplinească scopurile lor biologice. Dar opiniile au aceste scopuri biologice tocmai datorită adevărului principiului (D), și nu din vreun alt motiv.

Permiteți-mi acum să revin la cele două teme pe care le lăsasem suspendate, la sfârșitul secțiunii precedente – validitatea și dorințele generale.

Pentru a aborda dificultatea relativă la validitate, să considerăm încă o dată sistemul decizional general uman. Până acum am prezentat acest sistem ca și cum ar consta numai din opinii și dorințe. Sistemul în cauză are însă nevoie, în mod cert, de o „procedură inferențială“, astfel încât să fie capabil să infereze, din opiniile existente, concluzii cu privire la mijloacele adecvate. Această procedură inferențială poate fi însă mai mult ori mai puțin eficientă în producerea unor asemenea inferențe. Dar oricât de eficientă sau ineficientă ar fi ea, scopul său va fi să

producă astfel de inferențe *in mod valid*, adică, pornind de la premise adevărate, să deriveze concluzii adevărate cu privire la mijloace. Acest lucru e necesar pentru că în istoria noastră biologică procedurile inferențiale au fost favorizate tocmai în măsura în care ele s-au dovedit bune pentru așa ceva – adică, bune să ne spună care acțiuni vor produce de fapt rezultate reușite.

Obiecția considerată în ultima secțiune era că principiul (D) nu avea îndreptățirea să definească adevărul în termenii validității, pentru că validitatea însăși depinde de adevăr. Dar o dată ce principiul (D) este încorporat în teoria teleologică, acest lucru nu mai poate constitui o obiecție reală. *Atât adevărul cât și validitatea* au fost deja biologizate, în sensul că ambele sunt acum explicate ca aspecte ale funcționării cu succes a sistemului decizional uman. Opiniile își joacă rolul în acest sistem prin faptul că sunt adevărate, iar inferențele își îndeplinesc rolul prin aceea că sunt valide.

Desigur, este încă prezent un element de circularitate. Procedurile noastre inferențiale pot avea ca scop numai validitatea, în măsura în care ele sunt alimentate, de obicei, cu premise adevărate: inferențele valide nu pot conduce la acțiuni reușite dacă premisele lor sunt false. Convers, curentul general al opiniilor nu poate avea ca scop decât adevărul, în măsura în care inferențele construite pornind de la ele sunt în general valide: opiniile adevărate nu pot conduce la acțiuni reușite dacă concluziile derivate din ele sunt nevalide. Dar astfel de scopuri întrepătrunse sunt o caracteristică familiară a sistemelor biologice. De exemplu, plămânii noștri au doar scopul de a oxigena sângele, deoarece inima îl va pune în mișcare, însă ea are numai scopul de a asigura circulația. Selecția naturală este pe deplin capabilă să proiecteze sisteme care să asigure obținerea unor rezultate prin funcționarea coordonată a diferitelor componente. Și tocmai în acest sens a fost proiectat sistemul nostru decizional, adică pentru a alege acțiunile reușite construind inferențe valide, bazate pe opinii adevărate.

Să ne ocupăm acum de dorințele generale. În conformitate cu teoria teleologică, condițiile de satisfacere a dorințelor sunt acelea pe care chiar dorințele au ales să le producă. Asta ne indică diverse explicații posibile despre motivul pentru care unele



dorințe trebuie să fie generale. În primul rând, există un posibil argument bazat pe selecția genetică naturală. Astfel, de exemplu, se poate susține că istoria biologică a omului a favorizat organisme a căror constituție le-a dirijat spre ingerarea de  $H_2O$  atunci când au avut nevoie, pe când organismele dintr-o lume geamănă ar fi putut fi favorizate să aibă stări care să le dirijeze spre căutarea de XYZ, atunci când le-ar fi lipsit. Acceptând teoria teleologică, asta ar însemna că dorința pe care o exprimăm prin „apă“ va fi satisfăcută de  $H_2O$ , în timp ce dorința echivalentă, din lumea geamănă, ar fi satisfăcută de XYZ <sup>18</sup>.

Un dezavantaj al acestui argument biologic este, însă, că se restrânge la dorințele cu bază genetică. O evaluare mai complexă a dorințelor cu caracter general ar putea să facă apel la *învățare* – în particular, la procesele de învățare implicate în însușirea unui limbaj. Pare plauzibil că antrenamentul lingvistic este un proces de selecție a stărilor cognitive ce ne vor permite să vorbim corect: în acest sens, multe din abilitățile noastre lingvistice vor fi dobândite, în primă instanță, pentru a ne putea conforma *obiceiurilor* comunității. În mod independent este plauzibil și faptul că numele proprii și termenii ce desemnează genuri naturale sunt utilizați de comunitățile umane pentru a se referi la *obiectul sau materia care constituie originea cauzală a utilizării numelui*, sau poate la *obiectul sau materia pe care experții le vor recunoaște ca satisfăcând numele*. Cumulând aceste două aspecte, urmează că scopul opiniilor și dorințelor, exprimat prin astfel de nume proprii sau termeni ce desemnează genuri naturale, va fi co-prezent cu, ori va da naștere la asemenea obiecte și materii. Și astfel, dată fiind teoria teleologică, înseamnă că aceste opinii și dorințe vor avea condiții generale de adevăr și satisfacere. (Într-un fel, această evaluare socio-lingvistică a generalității asimilează numele proprii și termenii genurilor naturale cu termenii indexiali, așa cum ne indică expresiile scrise mai sus cu *italice*. Remarcăți însă că această indexialitate este, ca să zicem așa, localizată în comunitate, mai degrabă decât în capetele indivizilor: ceea ce face ca un concept precum „aluminu“ să stea pentru *substanța din fața noastră* este tocmai scopul conceptului respectiv – anume, acela de a permite indivizilor să vorbească corect atunci când aplică termenul la substanța localizată; iar

acest scop nu trebuie să fie reflectat de nimic din ceea ce s-ar afla în capul individului respectiv. De aceea, referenții numelui propriu și ai conceptelor ce desemnează genuri naturale nu se modifică la fel de ușor precum fac oamenii excursii în acea lume geamănă).

## 6. FALSITATEA FUNCȚIONALĂ

În această ultimă secțiune vreau să arăt în ce fel se atenuază una din obiecțiile comune la adresa teoriei teleologice a reprezentării, o dată ce recunoaștem că teoria teleologică mai degrabă încorporează evaluarea (D) a condițiilor de adevăr în termenii garanției reușitei – prezentată mai înainte – decât s-o înlocuiască. Aceasta este obiecția privind faptul că anumite opinii au scopuri biologice, care le pot solicita prezența chiar și atunci când sunt false (contrar pretenției că ar fi posibil să analizăm condițiile de adevăr ca circumstanțe în care opiniile sunt presupuse biologice prezente)<sup>19</sup>.

Să considerăm, de exemplu, opinia că nu ați fi pe cale de a fi răniți în cursul unei încleștări violente, iminentă și inevitabilă. Se poate susține că selecția naturală ne-a lăsat moștenire, ca dispoziție înăscută, să ne formăm această opinie – chiar și în cazurile în care ea ar merge împotriva evidenței –, pentru a se asigura că vom fi neclintii în bătălii. Dar, în conformitate cu teoria teleologică, de aici pare să urmeze că în condiția de adevăr a acestei opinii vor fi incluse o mulțime de circumstanțe în care noi *vom fi* răniți (din moment ce se presupune că noi suntem biologic determinați să avem opinia în cauză). Însă asta arată ca o *reductio* a teoriei teleologice. Căci prin ipoteză, condiția de adevăr a opiniei este aceea că noi *nu vom fi* răniți.

Exemplele de felul acesta sunt interesante, însă nu cred că sunt suficiente pentru a discredita teoria teleologică. Pentru a înțelege de ce nu, dați-mi voie să fac distincție între două tipuri diferite de scopuri biologice pe care le pot sluji opiniile. Le voi numi scopuri *normale* și scopuri *speciale*. Scopurile normale

implică contribuția opiniilor la sistemul decizional uman, așa cum a fost discutată în secțiunea precedentă. Scopurile speciale presupun contribuția opiniilor la alte nevoi biologice, în modalități care vor fi explicate pe scurt.

Teoria teleologică a reprezentării trebuie să fie înțeleasă ca referindu-se în mod specific la scopurile normale ale opiniilor. Să ne reamintim că opiniile contribuie la procesul decizional uman prin indicarea mijloacelor care vor duce la obținerea oricăror rezultate *dorite* în mod curent. Pentru a sluji *acest* scop, opinia din exemplul anterior va trebui, totuși, să fie adevărată, chiar dacă există alte scopuri speciale care i-ar cere să fie falsă. În acest fel, dovedind că noi ne concentrăm în mod specific asupra scopurilor normale ale opiniilor, acest exemplu încetează să mai fie o problemă pentru teoria teleologică.

Dacă nu sunteți convinși că opinia referitoare la invulnerabilitate trebuie să fie adevărată, pentru a-și realiza funcția sa normală, haideți să considerăm, de exemplu, cazul lui Gigi Lașu, care ar prefera să rămână neatins mai degrabă decât să câștige bătlăia. Dacă Gigi crede că nu va fi rănit – deși, de fapt, va fi –, atunci măcar acest lucru ar putea să-l împingă chiar și pe el să intre în luptă, și să sfârșească astfel prin a căpăta rana pe care era așa de îndârjit s-o evite. Dar atunci Gigi nu va obține ceea ce dorește – adică, mai presus de orice, să rămână nevătămat. În consecință, dacă ar fi ca Gigi să fie sigur de satisfacerea dorinței sale de a rămâne nevătămat, atunci opinia lui că nu va fi rănit trebuie să fie adevărată, întocmai așa cum cerea teoria teleologică invocată până acum.

Desigur, Gigi are dorințe întrucâtva nefericite. În particular, dorințele lui sunt nepotrivite din punct de vedere biologic, în sensul că satisfacerea dorințelor sale nu corespunde promovării șanselor sale generale de supraviețuire și reproducere. Acesta e punctul în care intră în joc scopurile *speciale* ale opiniilor. Importanța acestor scopuri speciale constă, efectiv, în aceea că ele evită rolul normal pe care îl au opiniile în satisfacerea dorințelor și, în schimb, garantează că agenții care manifestă dorințe biologic inadecvate nu vor sfârși prin a săvârși acțiuni biologic inadecvate. Lașii constituie un caz remarcabil. Dorințele lor nefericite denotă că sunt gata să sfârșească prin a fugi de

luptă, numai pentru a evita o vătămare. În consecință, pentru a-i proteja împotriva primejdiei biologice a unor asemenea consecințe, selecția naturală îi predispune să creadă că sunt invulnerabili – chiar dacă evidența nu garantează asemenea opinie –, oprindu-i astfel să realizeze acele acțiuni care le-ar satisface, de fapt, dorințele.

Ar putea să pară derutant că selecția naturală ar fi în stare să ofere unor opinii două scopuri diferite. La urma urmelor, este de presupus că selecția naturală proiectează sistemele biologice pentru un singur scop final – anume, moștenirea genelor. Așadar, de ce nu ar avea opiniile simplul scop de a asigura o asemenea moștenire genetică?

Răspunsul este, o dată în plus, legat de natura sistemului decizional uman. Luați aminte că acest sistem funcționează alegând întotdeauna acea acțiune care pare cea mai potrivită să asigure moștenirea genelor. El alege, mai degrabă, acea acțiune care pare cea mai potrivită să satisfacă dorințele existente. N-ar fi imposibil să ne imaginăm un sistem biologic de primul tip, care să urmărească întotdeauna și în mod direct moștenirea genelor. Dar pare evident că limitele capacităților noastre cognitive ne-ar împiedica să facem lucrurile în acest fel. Noi tindem, în schimb, pe termen relativ scurt la asemenea țeluri precum căldura, sexul, înghețata de ciocolată sau audierea înregistrărilor cu Chuck Berry.

În general, asemenea țeluri pe termen scurt se corelează rezonabil cu reușita biologică, acesta fiind, fără îndoială, motivul pentru care dorințele noastre înnăscute, ca și căile noastre de a dobândi dorințe ne-înnăscute, au evoluat așa cum au făcut-o. Dar satisfacerea dorințelor noastre nu va coincide întotdeauna cu reușita biologică (nu orice act sexual duce, ori poate să ducă, la reproducere). Iar asta înseamnă, în consecință, că există anumite riscuri biologice subsecvente modului nostru de a face diverse lucruri. Se poate totuși întâmpla ca unele din aceste riscuri să fie sub-produse inevitabile ale sistemului nostru decizional bazat pe dorințe: de exemplu, poate că e inevitabil ca oamenii să aibă dorințe extrem de puternice de a evita rănirile – și de aceea, *inevitabil* ca în anumite circumstanțe asta să-i facă să acționeze împotriva intereselor lor biologice. În acest fel s-ar ajunge la situația în care selecția naturală ar *interfera* cu operația normală a sistemului decizional, furnizându-ne opinii care să ne determine să acționăm în maniere ce ar frustra propriile noastre dorințe, dar ne-ar satisface nevoile biologice.

Permiteți-mi acum să rezum argumentul prezentat în această ultimă secțiune: anumite opinii au *într-adevăr unele* scopuri biologice, care le solicită să fie false față de dorințe. Totuși, asta nu invalidează teoria teleologică a reprezentării. Căci putem să interpretăm teoria teleologică în sensul că ea s-ar concentra în mod specific asupra scopurilor *normale* ale opiniilor – anume, acelea de a garanta satisfacerea dorințelor. Iar aceste scopuri normale nu cer niciodată opiniilor să fie false <sup>20</sup>.

## NOTE

<sup>1</sup> Vezi Dennett (1969, Cap. 9), (1987, Cap. 8); Fodor (1984); Millikan (1984), (1986); Papineau (1984), (1986), (1987); McGinn (1989, Cap. 2).

<sup>2</sup> Vezi Ramsey (1927, p. 159); Putnam (1978); Appiah (1986); Mellor (1988). Vezi, de asemenea, Stalnaker (1984, în special pp. 15–18), care susține, așa cum voi face și eu, că abordarea în termenii garanției succesului este parte – dar numai o parte – a adevărului despre adevăr.

<sup>3</sup> Field (1972) susține că fizicalismul nu permite relații semantice primitive. Ideea aceasta nu este, totuși, specifică fizicalismului.

<sup>4</sup> Vezi Papineau (1984), în special Secțiunea IX. Vezi, de asemenea, Field (1978) și McGinn (1982).

<sup>5</sup> Este, oare, menționarea adevărului *esențială* în astfel de explicații ale succesului? Există o vastă literatură asupra temei dacă explicațiile precum (C) pot sau nu pot fi înlocuite cu explicații bi-stadiale care (a) să explice mijlocul *s* prin opinia și dorința relevantă, și apoi (b) să explice succesul prin faptul că se produce *t*, și care, de aceea, nu fac nici o mențiune explicită asupra adevărului ca atare. (Vezi Loar, 1981, Devitt, 1984, Field, 1986). Cu toate acestea, dezbateră respectivă nu prezintă o relevanță imediată pentru preocupările mele actuale. Chiar dacă menționarea adevărului nu este *esențială* pentru explicarea succesului în orice ocazie particulară, ea poate fi totuși *revelatoare* pentru considerarea explicației succesului ca instanță a modelului general, potrivit căruia opiniile adevărate produc acțiuni reușite. În raport cu preocupările mele, întrebarea importantă nu este aceea dacă am putea să ne descurcăm cumva fără adevăr, în cazul că ar trebui s-o facem, ci, mai degrabă, ce rol joacă adevărul în gândirea noastră, dat fiind că nu suntem nevoiți să ne descurcăm fără el.

<sup>6</sup> Argumentul în sprijinul acestui principiu îl datorez lui Horwich (1987).

<sup>7</sup> Cf. Mellor (1988, p. 86).

<sup>8</sup> Acest aspect mi-a fost semnalat de Paul Horwich.

<sup>9</sup> Acest răspuns la obiecția privitoare la incertitudine mi-a fost sugerat de Hugh Mellor.

<sup>10</sup> Ar putea încă persista senzația că există unele opinii care n-ar fi relevante, nici chiar contrafactual, pentru succesul unei acțiuni. Ce s-ar putea spune despre opinia că *nu există agenți*, sau despre opinia că *toate acțiunile mele sunt sortite eșecului?* (Aceste exemple mi-au fost oferite de David Owens și, respectiv, David Sanford). În acest punct trebuie să apelăm la ideea de *compoziționalitate* a opiniilor. Am putea recunoaște că opiniile sunt făcute din componente („concepte“), și să căutăm explicația semnificației reprezentationale a acestor componente în termenii contribuției lor sistematice la condițiile de adevăr ale opiniilor în care intervin – adică, în termenii contribuției lor sistematice la condițiile ce garantează succesul acțiunilor bazate pe acele opinii. În acest fel putem spera să fixăm semnificația reprezentatională a unor asemenea concepte în termenii contribuției lor la opiniile care pot fi relevante pentru acțiune, iar apoi să folosim acele valori reprezentationale spre a construi condiții de adevăr pentru asemenea opinii speciale, ca și cum nu ar putea fi relevante pentru acțiune.

<sup>11</sup> Vorbesc despre „opinia „ mai degrabă decât de „opinia că p“, astfel încât să fie clar că nu presupun conținutul reprezentational în explicarea ei.

<sup>12</sup> Ce se poate spune despre îngrijorarea că extensionalitatea lui (E) și (F) va permite existența unor instanțe fără conținut specificat? Astfel: opinia că zăpada este albă e adevărată dacă și numai dacă iarba este verde. Sau, echivalent: opinia că zăpada este albă va genera acțiuni reușite dacă și numai dacă iarba este verde. Nu e aici locul să urmărim această temă, dar aş vrea să sper că acestei îngrijorări i se va răspunde imediat ce analiza reprezentării va fi fost extinsă astfel încât să poată stăpâni compoziționalitatea opiniei. (Cf. nota 10).

<sup>13</sup> Dacă ar fi să privim instanțele lui (F) pur și simplu ca truisme despre adevăr, neexplicate de o analiză reductivă precum (D), atunci generalizarea ne-ar oferi numai o expresie nelămuritoare de forma „Orice opinie va fi adevărată numai în cazul în care ea este adevărată“.

<sup>14</sup> Acest element de corespondență distinge abordarea oferită aici de teoriile „necotabile“ (*disquotational*) ale adevărului, de felul celor apărute de Quine (1970), Leeds (1978) sau Horwich (1982). Teoriile necotabile sunt în acord cu teza redundanței, conform căreia nu există nimic care să poată face adevărată o opinie dată, mai degrabă decât o altă opinie p (unde p este orice enunț cu același conținut ca al acelei opinii), dar atunci neagă faptul că ar mai fi nevoie de o noțiune corespondentă a conținutului. Pentru o discuție detaliată a dificultăților cu care se confruntă asemenea teorii, vezi Field (1986).

<sup>15</sup> Cf. Dummett (1974).

<sup>16</sup> Un corolar ar fi acela că dacă nu am fi niciodată interesați în explicarea finalităților, ci numai a mijloacelor, atunci nu am pierde nimic prin adoptarea unei noțiuni strict „înguste“, sau chiar „sintactice“, de conținut al unei opinii. Remarcați, însă, că această observație se aplică numai dacă interesele noastre explicative se restrâng exclusiv la mijloace „de bază“, la lucruri pe care agentul le face potrivit voinței sale, dar nu se extind la mijloace care nu sunt de bază, precum acțiunile t, care sunt mijloace pentru alte rezultate, însă pe care agentul le efectuează făcând ceva diferit, s. Chiar dacă o astfel de t non-bazăică poate fi privită ca un mijloc pentru un rezultat ulterior, vom avea încă nevoie de implicarea conținutului reprezentational al opiniei că s va produce t, pentru a explica în ce fel a reușit agentul făcând în primul rând t.

<sup>17</sup> Vezi Papineau (1987, Cap. 4.2). Vezi, de asemenea, Dretske (1988), care nu numai că subliniază rolul învățării, dar susține că învățarea respectivă este esențială pentru tipul de teleologie implicat în reprezentările mentale pe deplin mature.

<sup>18</sup> Acest tip de argument în favoarea conținuturilor generale este explicat în McGinn (1989, pp. 157–9). În legătură cu originea exemplului H<sub>2</sub>O – XYZ, vezi Putnam (1975).

<sup>19</sup> Ned Block a susținut această obiecție în fața mea. Vezi Block (1982), p. 5, pentru o observație înrudită.

<sup>20</sup> Aș dori să mulțumesc lui Tom Baldwin, Ned Block, Paul Horwich, Hugh Mellor, Philip Pettit, Huw Price și Jamie White pentru că au discutat cu mine aceste teme.

Traducere Gabriel NAGĂȚ

## BIBLIOGRAFIE

- Appiah, A. 1986: "Truth Conditions: A Causal Theory" în J. Butterfield (ed.) *Language, Mind and Logic*, Cambridge University Press.
- Burge, T. 1979: "Individualism and the Mental", în *Midwest Studies in Philosophy*, 4.
- Burge, T. 1982: "Other Bodies", în Woodfield (ed.) (1982)
- Dennett, D. 1969: *Content and Consciousness*, Routledge and Kegan Paul.
- Dennett, D. 1987: *The Intentional Stance*, MIT Press.
- Devitt, M. 1984: *Realism and Truth*, Basil Blackwell.
- Dretske, F. 1986: "Misrepresentation", în R. Bogdan (ed.) *Belief*, Clarendon Press.
- Dretske, F. 1988: *Explaining Behaviour*, MIT Press.
- Dummett, M. 1974: *The Justification of Deduction*, British Academy Lecture, Oxford University Press.
- Evans, G. 1982: *The Varieties of Reference*, Clarendon Press.
- Field, H. 1972: "Tarski's Theory of Truth", *Journal of Philosophy*, 69.
- Field, H. 1978: "Mental Representation", *Erkenntnis*, 13.
- Field, H. 1986: "The Deflationary Conception of Truth", în G. MacDonald C. Wright (eds.) *Fact, Science and Morality*, Basil Blackwell.
- Fodor, J. 1984: "Semantics Wisconsin Style", *Synthese*, 59.
- Fodor, J. 1987: *Psychosemantics*, MIT Press.
- Horwich, P. 1982: "Three Forms of Realism", *Synthese*, 51.
- Horwich, P. 1987: "Pure Truth", lucrare nepublicată (1990).
- Leeds, S. 1987: "Theories of Reference and Truth", *Erkenntnis*, 13.
- Loar, B. 1982: *Mind and Meaning*, Cambridge University Press.
- McDowell, J. 1986: "Singular Thought and the Extent of Inner Space", în P. Pettit și J. McDowell (eds.) *Subject, Thought and Context*, Clarendon Press.
- McGinn, C. 1982: "The Structure of Content", în Woodfield (ed.) (1982).

- McGinn, C. 1989: *Mental Content*, Basil Blackwell.
- Mellor, D.H. 1988: "I and Now", în *Proceedings of the Aristotelian Society*, 89.
- Millikan, R. 1984: *Language, Thought, and Other Biological Categories*, MIT Press.
- Millikan, R. 1986: "Thought without Laws: Cognitive Science with Content", în *Philosophical Review*, 95.
- Neander, K. 1984: *Abnormal Psychobiology*, Ph. D. dissertation, La Trobe University.
- Papineau, D. 1984: "Representation and Explanation", în *Philosophy of Science*, 51.
- Papineau, D. 1986: "Semantic Reduction and Reference", în J. Butterfield (ed.) *Language, Mind and Logic*, Cambridge University Press.
- Papineau, D. 1987: *Reality and Representation*, Basil Blackwell.
- Putnam, H. 1975: "The Meaning of Meaning", în *Mind, Language and Reality*, Cambridge University Press.
- Putnam, H. 1978: "Reference and Understanding", în *Meaning and the Moral Sciences*, Routledge and Kegan Paul.
- Ramsey, F. 1927: "Facts and Propositions", în *Aristotelian Society Supplementary Volume*, 7.
- Stalnaker, R. 1984: *Inquiry*, MIT Press.
- Stich, S. 1982: "Dennett on Intentional Systems", în J. Biro și R. Shahan (eds.) *Mind Brain and Function*, Harvester.
- Quine, W. 1970: *Philosophy of Logic*, Prentice-Hall.
- Woodfield, A. (ed.) 1982: *Thought and Object*, Clarendon Press.
- Wright, L. 1973: "Functions", în *Philosophical Review*, 82.



# CAUZALITATEA MENTALĂ

*Tim CRANE*

(University College, Londra)

## I

**O dificultate pentru fizicalism.** Acest articol se referă la o dificultate situată în chiar centrul teoriilor fizicaliste contemporane asupra minții. Pe de o parte, fizicalismul a apărut inițial din nevoia de a explica locul cauzării mentale în lumea fizică. Pe de altă parte, în ultima vreme fizicaliștii au ajuns să considere explicarea cauzării mentale una din problemele lor majore. Dar cum este posibil acest lucru? Cum este posibil ca teoriile fizicaliste să aibă încă o problemă explicând ceva pe care fizicalismul lor intenționa să-l explice de la început? Dacă fizicalismul urmărește să fie o explicație a cauzării mentale, atunci de ce s-ar confrunta el încă cu problema cauzării mentale?

Lămurirea acestei probleme va pune în lumină atât dezbaterea recentă asupra cauzării mentale cât și fizicalismul însuși. Putem face o distincție clară între acele forme de fizicalism care identifică elementele mentale cu cele fizice și cele care pretind că există o oarecare relație mai slabă de „constituire” între mental și fizic. Cea de-a doua perspectivă este astăzi versiunea ortodoxă a fizicalismului. Am să argumentez că problema cauzării mentale constituie o problemă numai pentru acest fizicalism ortodox, nu și pentru teoriile identității. Prin ea însăși, aceasta nu este în mod particular o pretenție neobișnuită. Dar am să argumentez, de asemenea, că adevărata lecție care poate fi învățată din dezbaterea cauzării mentale este aceea că fizicalismul ortodox este fie instabil, fie nemotivat. Este instabil pentru că (spre deosebire de teoriile identității) nu poate reconcilia cauzarea mentală cu celelalte presupoziii fizicaliste. Este nemotivat, pentru că în încercarea de a rezolva problema cauzării mentale fizicalismul abandonează într-o manieră caracteristică una (sau mai multe) dintre presuposițiile care fac parte din motivarea inițială a fizicalismului.

Pentru a întemeia această pretenție trebuie să explic (a) natura argumentelor care susțin fizicalismul, (b) problema cauzării mentale și (c) soluțiile standard ale problemei. Aceste trei obiective vor constitui substanța acestui articol. Am să încep cu câteva considerații preliminare despre fizicalism.

## II

**Fizicalismul.** Prin „fizicalism“ înțeleg aici o teorie care pretinde că mentalul este esențialmente fizic. Îi urmez pe fizicaliști înțelegând prin „fizic“ tot ceea ce constituie obiect de studiu al fizicii sau al științei fizice. Deci, aici intră particulari fizici – cum ar fi atomii sau cuarcii – și proprietăți fizice – cum ar fi mase specifice sau viteze – precum și entități din orice altă categorie ontologică atâta vreme cât ele cad în domeniul fizicii. (Exista o întrebare importantă referitoare la ce trebuie să fie „fizica“, dar această întrebare trebuie să o ignor aici). Fizicalismul este astfel o teorie relativ a *posteriori*, al cărei conținut și justificare se bazează pe descoperirile empirice ale fizicii.

În literatura fizicalistă timpurie, expresia „este“ din propoziția „mentalul este esențialmente fizic“ era înțeleasă ca o identitate strictă. De curând însă, fizicaliștii înclină către înțelegerea lui „este“ ca fiind ceva mai apropiat de o relație de constituire. În acest sens, a spune că totul este fizic înseamnă a spune că totul este fie o entitate fizică, fie este constituit de, sau compus din entități fizice. Acest tip de fizicalism acceptă că există obiecte care nu sunt fizice – dar acestea sunt constituite în mod exhaustiv sau compuse din obiecte fizice.

Noțiunea de *constituire* sau *compunere* este o noțiune care se aplică cel mai clar la substanțe particulare – de pildă, pentru a descrie relația dintre o substanță și părțile ei. Cu toate acestea, de vreme ce fizicaliștii sunt interesați în primul rând de relația dintre *proprietăți* mentale și fizice, ei vor trebui să-și exprime fizicalismul în termenii unei noțiuni de constituire / compunere pentru proprietăți analoagă cu noțiunea de constituire / compunere pentru particulari.<sup>1</sup> Cum încercările de a face acest lucru se află într-un stadiu incipient și nu mi-am propus aici să le discut în detaliu, am să folosesc expresia „teoria constituirii“ (constitution

theory) doar ca o etichetă pentru acest tip de fizicalism. Folosind această etichetă nu vreau să sugerez însă că există o teorie complet elaborată pe care ea o desemnează.

Fizicaliștii au folosit de asemenea ideea de superveniență pentru a exprima relația dintre proprietăți mentale și fizice. Dar, în pofida sofisticării diferitelor încercări de a defini o relație adecvată pentru raportul mental-fizic, s-a ajuns la recunoașterea – chiar de către mulți fizicaliști – faptului că noțiunea de superveniență nu este suficient de tare pentru a caracteriza fizicalismul<sup>2</sup>. Din acest motiv, indiferent de meritele sau neajunsurile supervenienței, am să ignor noțiunea în cea mai mare parte a articolului.

Prin urmare, cele două mari tipuri de fizicalism sunt teoriile identității și teoriile constituirii. Prefer să clasific teoriile fizicaliste în acest fel, mai degrabă decât în termenii unei distincții mult mai uzuale între fizicalismul „reducționist” și „nonreducționist”, pentru două motive. În primul rând, conceptul de reducere este el însuși destul de controversat: există mai multe încercări de a explica ce este de fapt reducerea și există chiar o dispută printre fizicaliști referitoare la faptul dacă poate exista cu adevărat un fizicalism nonreducționist<sup>3</sup>. În al doilea rând, așa cum va reieși mai târziu, clasificarea teoriilor fizicaliste în acest fel separă teoriile fizicaliste care întâmpină dificultăți în problema cauzării mentale de cele care nu întâmpină astfel de dificultăți. Conform perspectivei standard din literatura curentă, cel care are astfel de probleme este fizicalismul „nonreducționist” și se întâmplă așa tocmai datorită faptului că este nonreducționist. Am să pun în discuție acest aspect: distincția dintre fizicalismul reducționist și cel nonreducționist nu ajută la formularea unui răspuns referitor la apariția problemei cauzării mentale.

### III

**Argumentele pentru fizicalism.** Pionierii teoriei identității. U.T. Place și J.J.C.Smart erau preocupați în primul rând să arate că teoria nu poate fi eliminată *a priori* pe temeiuri semantice sau conceptuale<sup>4</sup>. Deși influent, acest argument are un caracter negativ și oferă puține temeiuri pentru a crede în fizicalism.

Singurul argument, explicit pozitiv, oferit de Smart se baza pe briciul lui Occam. Dar acesta nu i-ar impresiona pe dualiștii cartezieni, dintre care mulți ar argumenta că postularea substanțelor mentale non-fizice este necesară pentru a explica anumite fenomene (e.g. experiența conștientă). Este nevoie de un argument independent.

Putem identifica trei argumente independente în favoarea fizicalismului care au dominat disputa. Primul este argumentul avansat de David Lewis și D. M. Armstrong. Așa cum îl prezintă Lewis, argumentul are două premise, o premisă *a priori* și una empirică. Premisa *a priori* este aceea că proprietățile mentale sunt definite de rolurile lor cauzale caracteristice: schemele lor sistematice de relații cauzale cu percepțiile, acțiunile și alte stări mentale. (Sau, cum spune Armstrong, „conceptul unei stări mentale este conceptul unei stări a persoanei, capabilă să manifeste un anumit tip de comportament“<sup>5</sup>.) Lewis apără această perspectivă „funcționalistă“ pe temeiul că ideea unei astfel de scheme sistematice este implicită în perspectiva simțului comun asupra atribuțiilor psihologice ale stărilor mentale.

A doua premisă, empirică, a lui Lewis este aceea că știința fizică este „adekvată din punct de vedere explicativ“: „există un corp unificat de teorii științifice de tipul celor pe care le acceptăm acum, care împreună oferă o explicație exhaustivă și adevărată a tuturor fenomenelor fizice. Ele sunt unificate în sensul că sunt cumulative: teoria care guvernează orice fenomen fizic este explicată de teorii care guvernează fenomenele din care acest fenomen este compus și modul în care el este compus din ele. Același lucru este valabil și pentru acestea din urmă, și așa mai departe, până la particulele sau câmpurile fundamentale, guvernate de câteva legi simple, concepute mai mult sau mai puțin în maniera fizicii teoretice de astăzi.“<sup>6</sup>

Ideea este, aici, că orice efect fizic trebuie să fie explicabil în termeni pur fizici – în termenii unor fenomene pur fizice. Din aceasta și din prima premisă decurge că proprietățile mentale sunt proprietăți fizice. Pentru că dacă ocupanții rolurilor cauzale mentale nu ar fi fizici, atunci, dată fiind a doua premisă, ei nu ar putea fi explicabili în termeni fizici și astfel nu ar putea avea nici un efect fizic. Dar, de vreme ce stă în natura lor să aibă efecte, ei trebuie să fie fizici. (Lewis nu elimină entitățile non-fizice, ci susține doar că ele nu pot avea efecte fizice și deci nu pot fi stări mentale).

Al doilea argument îl voi numi argumentul *supradeterminării*. El a fost susținut în ultimii ani de James Hopkins, Christopher Peacocke, David Papineau și alții<sup>7</sup>. Spre deosebire de argumentul lui Lewis, acest argument are în vedere mai degrabă particulari mentali (simboluri) decât proprietăți mentale (tipuri). El poate fi exprimat în felul următor. Să presupunem că un efect fizic simbolic E are o cauză mentală simbolică M și că toate efectele fizice au cauze în întregime fizice. De aici decurge că E are o cauză simbolică complet fizică, să zicem P. Deci, fie M și P îl determină pe E în mod independent – adică îl supradetermină pe E – fie M este identic cu P. Dar cum nu există supradeterminare, M și P sunt identice.

Trăsătura distinctivă a acestui argument o constituie negarea supradeterminării. Dar este aceasta atât de ușor de legat? Ne putem imagina cu ușurință cazuri aparent posibile în care supradeterminarea are loc. Să presupunem că un om este ucis fiind simultan împușcat și înjunghiat. Atât înjunghierea cât și împușcarea l-ar ucide, dar fiecare dintre aceste acțiuni ar avea același efect chiar dacă cealaltă nu s-ar petrece. Este oare acest gen de scenariu chiar imposibil?

Opiniile sunt împărțite în privința faptului dacă supradeterminarea nu poate avea loc sau doar nu are loc. Jaegwon Kim o consideră „absurdă”. Christopher Peacocke este mai precaut: el spune că posibilitatea unei astfel de supradeterminări este un lucru „pe care noi îl considerăm de obicei ca fiind fals și nu este clar de ce ar trebui să ne schimbăm părerea”<sup>8</sup>. Desigur, există o problemă specială în legătură cu supradeterminarea pentru analizele contrafactice ale cauzării, dar dacă dăm de-o parte acest aspect, perspectiva standard printre filosofii minții pare să fie aceea că supradeterminarea nu poate fi o trăsătură comună a lumii actuale. Așa cum spune Stephen Schiffer, „este greu de crezut că Dumnezeu este un inginer atât de prost”<sup>9</sup>. Am să mă întorc la supradeterminare.

Ultimul argument pe care am să-l discut este celebrul argument al lui Davidson, din „Mental Events”, care susține și el identitatea dintre evenimente mentale și fizice<sup>10</sup>. Davidson argumentează în felul următor: evenimentele mentale interacționează cu evenimentele fizice; atunci când două evenimente sunt

legate în calitatea de cauză și efect, ele au descripții sub care instanțiază legi „stricte“ ale naturii; și nu există nici o lege psiho-fizică strictă. Davidson reconciliază acest aparent conflict trăgând concluzia că toate evenimentele mentale sunt evenimente fizice. Căci, dacă un eveniment mental și unul fizic interacționează cauzal, ele au descripții sub care instanțiază legi stricte. Dar aceste legi trebuie să fie fizice de vreme ce nu există legi psiho-fizice stricte. Cu toate acestea, pentru a cădea sub o lege fizică, evenimentul mental trebuie să aibă o descripție fizică, iar dacă un eveniment are o descripție fizică atunci este un eveniment mental. De aceea toate evenimentele mentale care interacționează cu evenimente fizice sunt ele însele evenimente fizice.

#### IV

**„Completitudinea“ fizicii.** Toate aceste argumente se folosesc de o presuposiție comună cu privire la structura cauzală a lumii fizice. Acest lucru apare cel mai clar în a doua premisă a argumentului lui Lewis: „adecvarea explicativă“ a fizicii. Ce înseamnă aceasta? Răspunsul depinde parțial de sensul în care luăm noțiunea de explicație. Dacă considerăm explicațiile ca fiind ceea ce ne oferă de fapt teoriile științifice, atunci premisa spune că fizica poate să dea o explicație pentru toate efectele fizice. Acest lucru pare neplauzibil chiar și pentru un fizicist – majoritatea fizicaliștilor de astăzi nu cred că doctrina are nevoie de ideea că fizica explică totul. Fizicalismul contemporan admite că există multe tipuri de explicații indispensabile, dar non-fizice, ale fenomenelor fizice.

Putem întări argumentul lui Lewis interpretând această premisă în termeni de cauzare. Înțeleasă în acest fel premisa spune că toate efectele fizice au cauze fizice adecvate pentru a determina (sau a stabili) posibilitatea tuturor celorlalte fenomene. Ideea este că prin stabilirea cauzelor fizice (și a legilor fizice) sunt stabilite toate efectele fizice. Acesta este principiul pe care David Papineau îl numește „completitudinea fizicii“: „Toate efectele fizice sunt determinate sau posibilitatea lor este determinată de [cauze] fizice anterioare potrivit legilor fizice.“<sup>11</sup>

Am să-l urmez pe Papineanu în ce privește înțelesul termenului „completitudine“. Ideea care stă în spatele acestui termen nu este că fizica este o știință completă care cuprinde tot. Ideea este mai degrabă aceea că, cauzele sunt în întregime satisfăcătoare pentru efecte fizice sau stabilesc posibilitatea lor: nu sunt *necesare* alte cauze pentru a determina efecte fizice. De aceea le numim „complete“. (De acum încolo am să ignor indeterminismul de vreme ce nu este în mod direct relevant pentru argumentele pe care le discut).

Argumentul supradeterminării folosește, de asemenea, ca premisă completitudinea fizicii. Și asta deoarece, probabil, motivul pentru a considera că un anumit efect fizic are o cauză în întregime fizică este că toate efectele fizice au cauze în întregime fizice. Am să consider că premisa conform căreia toate efectele fizice au cauze în întregime fizice este în mare măsură aceeași cu premisa lui Lewis – completitudinea fizicii.

Este poate mai puțin evident faptul că, completitudinea fizicii reprezintă o premisă în argumentul lui Davidson. Dar, de fapt este o simplă consecință atât a caracterului nomologic al cauzării cât și al anomismului mentalului. Fie un efect fizic E care este produsul unei cauze C. Caracterul nomologic al cauzării spune că C și E au descripții sub care instanțiază o lege strictă. Dar anomismul mentalului spune că această lege strictă nu poate fi o lege mentală de nici un fel – deci trebuie să fie o lege fizică, iar C trebuie să fie fizică. Prin urmare, orice efect fizic trebuie să aibă o cauză în întregime fizică – din nou completitudinea fizicii.

## V

**Forma generală a argumentelor în favoarea fizicalismului.** După ce am izolat această presuposiție centrală în toate argumentele favorabile fizicalismului, forma generală a argumentelor ar trebui să fie de acum evidentă: reconcilierea cauzării mentale cu completitudinea fizicii prin identificarea elementelor

mentale cu elementele fizicii. Argumentul lui Lewis face acest lucru atribuind stărilor mentale roluri cauzale tipice și argumentând că aceste roluri cauzale sunt ocupate de stări fizice. Argumentul supradeterminării identifică datele mentale cu datele fizice pentru a bloca supradeterminarea prin cauze distincte – din nou, dată fiind completitudinea fizicii. Iar Davidson identifică evenimentele fizice și mentale pentru a face cauzarea mentală consistentă cu negarea legilor psiho-fizice și caracterul nomologic al cauzării, care împreună determină completitudinea fizicii<sup>12</sup>.

Există, desigur diferențe evidente între aceste argumente. Argumentul lui Lewis privește tipurile sau proprietățile, celelalte două privesc simbolurile sau particularii. Și, în timp ce Lewis, Armstrong și apărătorii argumentului supradeterminării consideră că argumentele lor sunt *a posteriori*, argumentul lui Davidson are un caracter relativ *a priori*. Ceea ce vreau să subliniez totuși este structura comună a argumentelor: încercarea de a explica cauzarea mentală, dată fiind completitudinea fizicii, prin identificarea cauzelor mentale cu cauze fizice. Astfel că nu pot fi de acord cu John Searle atunci când susține că: „una din presuposițiile împărtășite de majoritatea dualiștilor tradiționali și de fiziciști este aceea că, prin recunoașterea realității și eficienței cauzale a mentalului trebuie să respingem orice relație de identitate între fenomenele mentale și creier”<sup>13</sup>. Dimpotrivă, fiziciștii identifică fenomenele mentale cu fenomene ale creierului tocmai pentru că vor să păstreze eficiența cauzală a mentalului. Se pare, deci, că dacă există epifenomene mentale, atunci nu există nici un temei – în afară de considerațiile vagi și neconcluzive care țin de briciul lui Occam – pentru a spune că ele sunt fizice.

În aceste argumente se ascunde o presuposiție importantă. Am să o numesc „omogenitatea” cauzării mentale și fizice. Adică, pentru ca argumentele să funcționeze, trebuie să existe un conflict între cauzarea mentală și completitudinea fizicii. Dar dacă este așa, atunci noțiunea de cauzare este aceeași noțiune aplicată fizicului și mentalului deopotrivă. Deci, formula generală a argumentelor în favoarea fizicalismului nu este aceea că dacă există cauzare mentală aceasta trebuie să fie ceva atât de straniu și *sui generis* încât trebuie, pe motive de parcimonie și simplitate teoretică, să o respingem. Pentru că răspunsul evident la aceasta este recunoașterea faptului că noțiunea *sui generis* de cauzare mentală este la fel de stranie (sau nu) ca aceea de cauzare fizică.



Nu există nici un conflict și astfel nu este nevoie de o teză de identitate – dacă noțiunile de cauzare folosite sunt atât de diferite. Cu alte cuvinte, argumentele pentru fizicalism trebuie să plece de la presupuziția că etichetele „mental“ și „fizic“ așa cum se aplică cauzării sunt într-adevăr epitete transferate – mentale sau fizice –, sunt relatele cauzării și nu cauzarea însăși.

## VI

**Fizicalismul ortodox.** Varianta fizicalismului, motivată de aceste argumente este într-o versiune sau alta) teoria identității. Este demn de observat totuși faptul că doar câțiva fizicaliști contemporani acceptă de fapt teoriile identității. Se consideră în general că aceste teorii sunt fie prea tari pentru a fi plauzibile, fie prea slabe pentru a fi explicative. În schimb, fizicaliștii sunt înclinați să susțină o versiune a teoriei constituirii menționată în paragraful II – teoria constituirii a devenit versiunea oficială a fizicalismului<sup>14</sup>. Cum apare această trecere de la identitate la constituire?

Cea mai importantă cauză este influența obiecției realizării variabile a lui Putnam la teoria identității de tip: pare nomologic posibil ca multe entități fizice diferite să poată fi toate în același tip de stare mentală. Deci, teoria identității de tip este mult prea tare pentru a fi plauzibilă empiric. Pe de altă parte, teoria identității simbolice pare prea slabă pentru a fi satisfăcătoare: cum se explică *de ce* aceste simboluri mentale sunt identice cu acest simboluri fizice? Se presupune că o soluție la problema minte-corp ar da un răspuns clarificator întrebării referitoare la relația dintre mental și fizic. Dar este greu de spus cum poate realiza acest lucru teoria identității simbolice (Obiecția ar putea să apară mai clar cu ajutorul unei analogii: a susține pur și simplu pretenția de „identitate simbolică“, că fiecare președinte al Statelor Unite este identic cu un anumit bărbat alb, poate cu greu fi considerată o explicație a faptului că toți președinții Statelor Unite au fost bărbați albi).

În istoria recentă a fizicalismului intervine aici noțiunea de superveniență: proprietățile fizice determină proprietățile mentale, dar nu sunt identice cu acestea. Totuși, superveniența nu este suficient de tare pentru a explica relația mental-fizic – deci mai este nevoie de ceva pentru a face din superveniență o versiune a fizicalismului. Acest „ceva“ este dat de ceea ce am numit „constituire“. (Alți autori folosesc alți termeni: David Charles și Philip Pettit folosesc termenul de „compunere“ iar într-o examinare recentă a problemei supervenienței, Terence Horgan introduce cu optimism neologismul „ultrasuperveniență“<sup>15</sup>).

Această evoluție a teoriilor – de la teoriile identității *via* superveniența, la teoriile constituirii poate fi cel mai bine apreciată ca o dezbatere internă între fizicaliști. Teoria constituirii capătă doar rareori o motivație independentă. Abordarea urmată de obicei de fizicaliști este: „cu toți știm că trebuie să fim fizicaliști – întrebarea este, care este forma acceptabilă de fizicalism?“<sup>16</sup> (Importanța acestui aspect va reieși la final).

În orice caz, nu se poate nega faptul că versiunea ortodoxă a fizicalismului este acum teoria constituirii: teoria conform căreia proprietățile mentale sunt constituite din proprietăți fizice. După cum am afirmat și în paragraful II, nu am de gând să discut în detaliu ce se înțelege prin „constituire“, dar, în parte, ideea este probabil aceea că proprietățile fizice constituie proprietăți mentale în măsura în care sunt instanțiate: Durerea mea de cap de *acum* este constituită din unele proprietăți fizice instanțiate în creierul meu. Nu există nici o legătură cu ideea că universalii mentale neinstanțiate, în măsura în care asemenea lucruri există, sunt constituite din universalii fizice neinstanțiate. Cu toate acestea, universalii neinstanțiate sunt totuși universalii, iar aceasta înseamnă că teoria constituirii este o teorie a „identității simbolice“ în sensul cel mai „slab“. Să presupunem că eu instanțiez proprietatea *durere* și că, de asemenea, instanțiez o anumită proprietate cerebrală, *B*, presupusă a constitui durerea. Există deci două entități complexe: instanțierea mea a proprietății *durere* și instanțierea mea a proprietății *B*. Deși unii filosofi (în special Kim) numesc astfel de entități complexe „evenimente“, eu le voi numi „fapte“ (ele mai pot fi numite și „stări de fapt“)<sup>17</sup>. Această teorie este o teorie a identității simbolice numai în sensul că individul care are proprietatea *durere* este același cu individul

care are o proprietate a creierului. Adică eu. De vreme ce proprietățile *durere* și *B* sunt distincte, faptele care le încorporează trebuie să fie de asemenea distincte. Faptul că eu sunt constituit din aceste fapte împreună este desigur adevărat, dar acest lucru are foarte puțin de-a face cu problema fizicalismului și nu poate decât să conducă la denumirea în mod eronat a acestui punct de vedere – teoria identității simbolice<sup>18</sup>.

Cu toate acestea, există o poziție, apărută de Graham și Cynthia MacDonald, care pretinde că identice nu sunt proprietățile însele, ci *instanțele* proprietăților mentale și fizice („instanțe-proprietate”)<sup>19</sup>. Ar putea să pară deci că această perspectivă nu se reduce doar la acea teorie slabă a „identității simbolice”.

Ce reprezintă această poziție depinde de felul în care înțelegem „instanțele-proprietate”. Cei doi MacDonald nu înțeleg prin acestea nici lucruri care au o proprietate (particulari) nici *tropi* („particularii abstracti” pe care unii îi consideră elementele de bază ale realității)<sup>20</sup>. Prin instanțe-proprietate ei înțeleg „evenimente” (oarecum asemănător cu sensul lui Kim) ale căror „componente constitutive” sunt obiecte, proprietăți și momente<sup>21</sup>. Deci se presupune că instanțele-proprietate sunt entități distincte de obiectele care au proprietăți și de proprietățile (universalile) pe care le instanțiază – chiar dacă aceste obiecte și proprietăți sunt „componente” ale instanțelor-proprietate.

Acum este greu de spus cum diferă instanțele-proprietate de fapte, după definiția de mai sus: lucrurile au proprietăți în anumite momente de timp. Aceasta pare să fie singura diferență. Pare plauzibil să spunem că faptul că *Simt o durere la momentul t* și faptul că *Am o stare a creierului B la momentul t* sunt unul și același fapt numai în cazul în care a *simți o durere* este aceeași proprietate ca *B*. Dar acest lucru nu este adevărat în privința instanțelor-proprietate ale celor doi MacDonald. Perspectiva lor este aceea că o instanță conține ca elemente „componente” proprietatea a *simți o durere* și proprietatea *B*. Asta vor să spună ei prin aceea că proprietățile mentale și fizice pot fi „instanțiate într-o singură instanță”<sup>22</sup>. Deși este corect să spunem că instanțele-proprietate mentale și fizice sunt *identice*, acest lucru ar putea să ne inducă în eroare – soții MacDonald vor să spună că o singură instanță-proprietate are drept „componente” o proprietate mentală și una fizică.

Totuși, nu este clar de ce cineva care susține teoria faptului *nu* poate susține că un fapt poate conține drept „componente” atât o proprietate mentală cât și una fizică. Și dacă lucrurile stau astfel, atunci nu prea avem de ales între „teoria faptului” și “teoria instanței-proprietate”, iar perspectiva celor doi MacDonald nu este decât „teoria faptului” formulată în alți termeni<sup>23</sup>. Din acest motiv, am să tratez teoria lor ca o versiune a „teoriei constituirii”. Ceea ce merită atenția în cadrul teoriei este ideea că proprietățile mentale și fizice sunt (într-un anumit fel) „unite”, dar nu identice.

Trecerea de la identitate la constituire (și la noțiunile înrudite) este privită uneori ca fiind inofensivă, fără consecințe ontologice importante. Într-o încercare recentă de a defini fizicismul, de pildă, Philip Pettit tratează identitatea ca fiind doar un caz special al „compunerii”<sup>24</sup>. Dar mie mi se pare că trecerea este foarte semnificativă, de vreme ce, odată acceptată, fizicaliștii trebuie să se confrunte cu problema cauzării mentale.

## VII

**Problema cauzării mentale.** Putem construi o explicație simplă a acestei probleme în felul următor. Cauzele au efecte în virtutea proprietăților lor. Dacă arunc cu o cărămidă în geam și geamul se sparge, acest lucru se va întâmpla datorită unor proprietăți specifice ale cărămidii. (Nu contează aici dacă acestea sunt luate ca proprietăți ale cărămidii sau ale unui eveniment – evenimentul constând din lovirea geamului cu cărămidă). Dar nu toate proprietățile unei cauze sunt responsabile pentru efectele ei. Spargerea geamului nu se datorează culorii cărămidii sau strălucirii ei sau „proprietăților relaționale”, cum ar fi faptul că este fabricată în Walthamstow, sau faptului că a fost aruncată exact la 4.05 pm.

Să presupunem că stările mentale particulare nu sunt identice cu stările fizice particulare. În aceste condiții ne putem întreba: sunt oare proprietățile mentale ale cauzelor responsabile pentru aceste efecte fizice? Se ridică astfel următoarea dilemă: dacă

proprietățile mentale sunt responsabile pentru aceste efecte și efectele nu sunt supradeterminate, cum se împacă acest lucru cu completitudinea fizicii? Conform completitudinii fizicii, cauzele fizice (adică *trăsăturile* fizice ale cauzelor) sunt suficiente pentru a determina efectele fizice. Deci, cum pot avea proprietățile mentale o răspundere cauzală?

Pe de altă parte, dacă proprietățile mentale ale cauzei *nu* sunt responsabile pentru efectele sale, atunci se pare că epifenomenalismul este adevărat: mentalul nu joacă nici un rol din punct de vedere cauzal. Și deci fizicalismul ortodox este fie inconsistent, fie epifenomenalist – aceasta este problema cauzării mentale pentru fiziciști.

Să observăm că acest argument presupune ceea ce am numit în paragraful V „omogenitatea” cauzării mentale și fizice. Dacă omogenitatea nu se susține, atunci problema dispare. Să mai observăm că datorită felului în care este pusă problema – „oare cauzele determină efectele în virtutea proprietăților lor mentale” – este clar că problema nu se pune pentru teoreticienii care susțin identitatea de tip.

Așa cum au observat unii autori, există de fapt două întrebări separate, referitoare la fenomenalism, implicate în această problemă<sup>25</sup>. Prima este aceea dacă stările sau evenimentele mentale particulare sunt cu adevărat cauze. A doua este aceea dacă stările sau evenimentele mentale sunt cauze în virtutea proprietăților lor mentale. S-ar putea susține că stările sau evenimentele mentale sunt cauze în virtutea proprietăților lor mentale – dacă, de exemplu, stările sau evenimentele mentale simbolice ar fi identice cu stări sau evenimente fizice simbolice. Am să cercetez astfel de poziții în secțiunea următoare.

Trebuie să facem totuși o distincție între problema cauzării mentale și problema „eficienței cauzale a conținutului”<sup>26</sup>. Problema referitoare la eficiența conținutului este problema modului în care conținuturile stărilor intenționale pot fi relevante pentru efectele acelor stări. Problema cauzării mentale se pune totuși pentru toate stările mentale, nu numai pentru acelea care au conținut intențional. Mai mult, problema referitoare la conținut nu este aceea că ar fi *mental* – o astfel de idee nu prea are sens, oricum. Să presupunem, de dragul argumentului, că totul ar fi

fizic și faptul că o stare posedă conținut intențional a fost redus la fapte pur fizice referitoare la starea respectivă și la relațiile sale cu mediul. În aceste condiții se poate ridica totuși întrebarea: efectele unei stări intenționale se datorează oare faptului că are acest conținut? Faptul că se poate pune această întrebare arată că, chiar dacă teoria „identității de tip“ ar fi adevărată, problema eficienței conținutului ar rămâne încă nerezolvată. În schimb, în felul în care am construit problema cauzării mentale, această dificultate nu mai apare dacă teoriile „identității de tip“ sunt presupuse ca adevărate. Și nu mai apare problema dacă există *explicații* mentale adecvate – explicații care să invoce stări sau evenimente mentale în explicarea comportamentului. Întrebarea este ce anume face ca aceste explicații să fie adevărate – în particular, sunt ele adevărate în virtutea faptului că stările sau evenimentele mentale se află printre cauzele care determină comportamentul?

Tocmai din acest motiv (printre altele) problema nu se rezolvă doar prin introducerea unui criteriu contrafactual al eficienței mentale. Nu este de ajuns să spunem că eficiența mentală este asigurată de adevărul unor contrafactuali, cum ar fi „dacă proprietatea mentală nu ar fi fost instanțiată, efectul fizic nu ar fi apărut“. Este bine-cunoscut faptul că astfel de contrafactuali pot fi adevărați fără ca antecedentul să desemneze o cauză a fenomenului în discuție. De exemplu, adevărul acestor contrafactuali este consistent cu proprietatea mentală și efectul fizic ca efecte ale unei cauze comune<sup>27</sup>.

Deși aici disting între cauzare și explicație, nu trebuie să spun prea multe despre ceea ce presupune disputa referitoare la natura cauzării, de vreme ce, în mare parte, ea nu depinde de nici o teorie specifică a cauzării. (O excepție importantă este teoria lui Davidson, pe care am să o discut în secțiunea următoare). Toți participanții la dispută pot presupune, de exemplu, că toate cauzele deterministe sunt necesare și suficiente (în circumstanțe) pentru efectele lor, sau că ele trebuie să sporească șansele efectelor lor – să facă astfel ca acele efecte să fie mai probabile decât ar fi fost altfel – sau că existența lor implică existența unei legi căreia i se supun.

Există totuși o presuposiție referitoare la cauzare care este esențială argumentului. Anume, aceea după care cauzele își

determină efectele în virtutea unora dintre proprietățile lor. Atunci când caracterizăm cauzele alegem anumite trăsături ale lor – „proprietăți eficiente cauzal” – în virtutea cărora ele au efectele pe care le au<sup>28</sup>. Nu trebuie să apară aici această presuposiție de vreme ce ea este acceptată de aproape toți participanții la dispută.

Dar cine sunt participanții? Și cine ar trebui să-și facă griji referitor la problema cauzării mentale?

## VIII

**A cui este problema?** Dezbaterea contemporană care se desfășoară în jurul problemei cauzării mentale a apărut în special prin criticismul monismului anomalistic al lui Davidson<sup>29</sup>. Criticii lui Davidson au ridicat de fapt tocmai întrebarea noastră – „oare evenimentele mentale își determină efectele în virtutea proprietăților lor mentale?” și au argumentat că, dacă Davidson susține că evenimentele mentale au efectele pe care le au în virtutea proprietăților lor fizice, atunci evenimentele mentale sunt redundante cauzal. Deci monismul anomalistic este fie inconsistent fie epifenomenalist.

Cu toate acestea, în pofida discutării pe larg a acestei probleme în literatura de specialitate, ea nu constituie o dificultate pentru Davidson. Și asta deoarece pentru Davidson, cauzarea este o relație între evenimente particulare indiferent de proprietățile pe care le au – sau, pentru a folosi stilul nominalist al lui Davidson, indiferent de modul în care sunt descrise. Așa cum spune Davidson însuși, „există evenimente care au cauze și efecte. Dată fiind perspectiva extensionalistă asupra relațiilor cauzale, nu are literalmente nici un sens (...) că vorbim de un eveniment care cauzează ceva în virtutea faptului că este mental, sau în virtutea proprietăților sale mentale, sau în virtutea descrierii lui într-un fel sau celălalt”<sup>30</sup>. Acesta nu este, după Davidson, epifenomenalism. Ar fi epifenomenalism dacă ar avea sens să spunem

că trăsăturile fizice ale evenimentelor mentale ar fi mai eficace decât mentalul. Dar nu putem spune acest lucru: pentru că este de asemenea „irelevant pentru eficiența cauzală a evenimentelor fizice faptul că ele pot fi descrise într-un vocabular fizic“<sup>31</sup>.

Ideea centrală aici este aceea ca Davidson respinge principiul menționat în ultima secțiune și anume: cauzele au efectele pe care le au în virtutea unora dintre proprietățile lor. El respinge acest principiu nu pur și simplu pentru că respinge proprietățile, ci pentru că susține despre cauzare că este o relație între particulari. Cu alte cuvinte, dacă unele proprietăți ale unei cauze fac ca ea să fie „mai eficace“ decât altele, atunci unele modalități de a descrie cauza respectivă sunt mai bune decât altele – unele selectează trăsăturile relevante cauzal, altele nu. Dar, conform teoriei lui Davidson, nu există nici o modalitate de a descrie o cauză ca fiind „mai eficace“ decât în alte descrieri. Toate descripțiile adevărate ale unei cauze descriu ceva eficient – altfel ele n-ar fi descripții adevărate ale cauzei.

Aceasta nu înseamnă că Davidson nu poate da un răspuns la întrebarea „oare cărămida a spart fereastra în virtutea masei sale sau a culorii?“ Dar el va considera această întrebare o cerere de a da o *explicație* cauzală informativă și nu o întrebare referitoare la proprietățile *eficiente* ale cauzei. Și așa cum am văzut în secțiunea precedentă, chestiunea explicației este în mare parte irelevantă pentru problema noastră.

Cu toate acestea, se poate spune că Davidson trebuie să se confrunte cu problema dacă, contrafactualii care au următoarea formă sunt adevărați: „dacă o anumită descripție mentală a cauzei n-ar fi fost adevărată, atunci efectul nu ar fi avut loc“. Și în acest fel Davidson ar putea să se angajeze, cel puțin față de o înțelegere a cauzelor ca avându-și efectele „în virtutea“ anumitor proprietăți: viz., faptul că  $y$  este  $P$  are loc în virtutea faptului că  $x$  este  $M$ ; înseamnă că dacă „ $M$ “ n-ar fi fost adevărat despre  $y$ , atunci „ $P$ “ n-ar fi fost adevărat despre  $y$  (unde „ $M$ “ și „ $P$ “ sunt predicate mentale și fizice). Totuși, chiar dacă Davidson ar accepta acest „în virtutea a“, de aici nu ar decurge o problemă în privința cauzării mentale, de vreme ce această pretenție de dependență contrafactuală nu intră în conflict cu completitudinea fizicii, așa cum o înțelege el<sup>32</sup>.



Dacă această interpretare a lui Davidson este corectă, ea ridică următoarea întrebare: de ce se consideră de obicei că Davidson se confruntă cu problema cauzării mentale? Eu cred că acest lucru se datorează, în parte, faptului că uneori ideea unei teorii a identității simbolice este definită în literatura filosofică independent de teoria cauzării a lui Davidson<sup>33</sup>. Concepția lui Kim despre „evenimente“ care cuprind proprietăți este de asemenea bine-cunoscută și acceptată, ca și legătura dintre proprietăți și cauzare; dacă aceste idei constituie fundalul unei lecturi a lucrării lui Davidson („Mental Events“), atunci pare natural să ridicăm problema cauzării mentale pentru Davidson. Este ca și cum oamenii ar gândi: „Davidson a arătat cum să fii fizicist fără să fii reduționist. Dar acum trebuie să decidem care „teorie a evenimentelor“ și care „teorie a cauzării“ să o acceptăm.“ Dar, dacă acceptăm nu numai concluzia lui Davidson, ci și argumentul prin care el o susține, nu ne mai putem întreba dacă să acceptăm sau nu teoria lui, referitoare la evenimente și cauzare.

Mult mai important este, totuși, faptul că există ceva foarte nesatisfăcător referitor la negarea (implicită) de către Davidson a tezei după care cauzele au efecte în virtutea unora dintre proprietățile lor. Scopul acestei teze este de a marca o distincție crucială între acele proprietăți a căror instanțiere produce o diferență causală (într-un caz particular) de acele instanțieri care nu produc o astfel de diferență<sup>34</sup>. Așa cum am arătat mai înainte, nu se poate spune că Davidson nu marchează această distincție. Dar el o va marca ca o distincție între explicații, și nu între proprietăți ale lucrurilor. Problema este că teoria lui Davidson pare să nu fie capabilă să răspundă la întrebarea de ce anumite explicații sunt mai bune decât altele prin invocarea *trăsăturilor* eficace ale realității. Deci, perspectiva mea este că deși Davidson nu întâmpină dificultăți în privința cauzării mentale, teoria sa are totuși probleme în altă parte.

Cu toate acestea, faptul că Davidson nu are probleme cu cauzarea mentală este motivul pentru care nu vreau să-i caracterizez pe aceia care se confruntă cu problema cauzării mentale, ca „fizicaliști non-reduționiști“ – de vreme ce Davidson însuși este un fizicist non-reduționist proeminent și totuși nu se confruntă cu această problemă. Cei care întâmpină dificultăți sunt cei care

susțin „teoria constituirii”: să luăm un fapt sau un eveniment mental care pare să fie cauza unui anumit fapt sau eveniment fizic. Dacă „teoria identității de tip” este falsă, completitudinea fizicii este adevărată și nu există supradeterminare, iar cauzarea depinde de proprietăți ale lucrurilor, atunci problema este clară: faptul mental nu poate fi cauza<sup>35</sup>.

Cea care are probleme cu cauzarea mentală este, prin urmare, „teoria constituirii”, oricum ar fi descrisă – în termeni de superveniență, compunere, realizare sau dependență. Deci, Jerry Fodor nu se putea înșela mai mult atunci când spunea că „superveniența minte-creier este cea mai bună idee pe care a avut-o cineva, vreodată, referitor la felul în care este posibilă cauzarea mentală”<sup>36</sup>. În afară de cazul limitativ al „identității de tip” – în care superveniența se susține trivial – teoria supervenienței /constituirii / realizării nu ne dă nici o idee despre cum este posibilă cauzarea mentală.

## IX

**Presupozițiile problemei.** La prima vedere deci, fizicalismul ortodox nu poate fi susținut pentru că nu oferă nici o soluție pentru a reconcilia cauzarea mentală cu completitudinea fizicii. Au existat, desigur, multe încercări de a realiza această reconciliere. Dar, se dovedește acum că toate aceste încercări sfârșesc prin a nega una din presuposițiile de bază ale fizicalismului.

Pentru a înțelege de ce se întâmplă acest lucru trebuie să vedem care sunt presuposițiile problemei cauzării mentale, așa cum am formulat-o mai sus. Acestea sunt următoarele:

- (A) Cauzele au efecte în virtutea (unora dintre) proprietățile lor.
- (B) Există cauzare mentală.
- (C) Completitudinea fizicii este adevărată.
- (D) Nu există supradeterminare.
- (E) Cauzarea mentală și fizică sunt „omogene”.

Mie mi se pare că singura cale de a reconcilia toate aceste presupoziiții este aceea de a susține „teoria identității de tip”: anume, de a identifica proprietățile mentale cu cele fizice. (Acest lucru ar trebui să fie deja evident datorită asemănării dintre forma generală a argumentelor care susțin teoria identității și presupoziițiile (A) – (E)).

În schimb, dacă respingem „teoria identității de tip”, ceea ce fac majoritatea participanților la dispută, problema poate fi rezolvată numai prin respingerea uneia sau mai multor presupoziiții (A) - (E). Așa cum am văzut în ultima secțiune, teoria lui Davidson o respinge pe (A). Ce se poate spune despre celelalte presupoziiții? Un epifenomenalist o va respinge pe (B). Dar aceasta este în mod sigur ultima presupoziiție pe care ar trebui să o respingem – faptul că mințile noastre ne face trupurile să se miște nu este o parte dintr-o teorie filosofică, ci un lucru pe care teoria ar trebui să-l explice.

Filosofii contemporani se gândesc numai rareori la posibilitatea de a respinge premisa (C), completitudinea fizicii. Unul din motive pare să fie acela că ei cred că respingerea completitudinii înseamnă angajarea față de dualismul cartezian<sup>37</sup>. Mie mi se pare că aceasta este o greșeală, pentru că ignoră posibilitatea unei poziții, conform căreia efectele fizice pot avea mai multe feluri de cauze, nici una dintre acestea cu trăsăturile pe care le atribuie Descartes minții și fără să fie toate fizice („fizic” în sensul precis de care are nevoie fizicalismul).

Conform completitudinii fizicii, fiecare efect fizic este determinat de cauze pur fizice, în acord cu legile fizice. Cu alte cuvinte, (ignorând indeterminismul) cauzele pur fizice sunt suficiente pentru apariția fiecărui efect fizic. Dar, orice altceva ar însemna „suficiente”, în cazul de față trebuie să însemne cel puțin „suficiente în circumstanțe” – chiar atunci când indeterminismul este în afara discuției. Aceasta este exact ideea familiară după care aprinderea unui chibrit (să zicem) este suficientă pentru aprinderea unui foc numai dacă există oxigen, material inflamabil și așa mai departe. La fel, prezența oxigenului este suficientă pentru aprinderea bățului de chibrit dacă acesta este folosit și dacă există material inflamabil – deci și aceasta este o cauză a aprinderii chibritului.

În mod asemănător, să presupunem că cineva aruncă o cărămidă într-un geam pentru că vrea să facă asta și geamul se sparge. Sunt de acord cu fiziciștii că, în aceste circumstanțe, starea creierului persoanei respective este suficientă pentru a-i pune în mișcare mușchii și pentru aruncarea cărămidii și pentru spargerea geamului. Dar, dacă circumstanțele includ și dorințele și opiniile persoanei în cauză, atunci și acestea vor fi suficiente, dată fiind schimbarea circumstanțelor, pentru spargerea geamului. Deci, oricare ar fi sensul în care cauzele fizice sunt suficiente pentru efectele lor, cauzele mentale pot face același lucru.

Dar asta înseamnă oare (așa cum susțin unii fiziciști) că există „goluri“ în lanțul cauzării fizice pe care opiniile și dorințele trebuie să le „umple“? Ce înseamnă că există astfel de goluri? Așa cum observa Tyler Burge, a gândi în acești termeni înseamnă „a considera cauzele mentale un model fizic – care dă un „brânci“ în plus efectului“<sup>38</sup>. Iar tendința multor fiziciști de a vorbi în termeni de cauze care necesită „forțe“ sau „vigoare“ arată că ei gândesc deja în termenii unui model de cauzare care face să fie greu de înțeles cauzarea mentală autonomă<sup>39</sup>. Acest model al cauzării poate să fie profund eronat: îmbrâncitul și lovitul sunt doar cazuri speciale de cauzare, nu noțiuni care să clarifice cauzarea. Același lucru este valabil și pentru ideea că „golurile“ trebuie să fie umplute.

Prin urmare, a respinge completitudinea fizicii nu înseamnă a respinge pretenția că, cauzele fizice sunt cauze suficiente pentru toate efectele fizice. Ci înseamnă a respinge pretenția că numai cauzele fizice sunt cauze suficiente pentru efectele fizice. A susține că doar cauzele fizice sunt cauze suficiente pentru efectele fizice nu implică prin ea însăși pretenția implauzibilă că în așa-zisele cazuri de cauzare mentală, dacă fenomenul mental n-ar fi existat, efectele fizice ar fi apărut totuși<sup>40</sup>.

Deci, conform poziției pe care o descriu, efectele fizice, ca orice alte efecte, pot avea mai multe cauze – unele dintre ele sunt mentale, altele sunt fizice. Această pretenție inofensivă nu intră în conflict cu legile fizicii și nici cu principii metodologice perfect legitime – de pildă, principiul că ar trebui să căutăm mecanismele care se ascund în spatele fenomenelor. Acest principiu metodologic este uneori invocat ca motivația principală care stă în

spatele completitudinii fizicii<sup>41</sup>. Dar, ar trebui accentuat că acceptarea acestui principiu (care este o maximă regulativă folositoare) nu cere acceptarea completitudinii fizicii (care este o pretenție empirică foarte tare). Este important să separăm completitudinea fizicii de acest principiu metodologic, pentru că altfel non-fizicalismul va primi, prin asociere, aceleași obiecții.

Așa cum respingerea completitudinii fizicii nu duce neapărat la dualismul cartezian, susținerea completitudinii fizicii nu implică, neapărat fizicalismul. Pentru că și (D) poate fi respinsă: acțiunile sunt supradeterminate de cauze mentale și fizice. Așa cum am văzut în paragraful III, supradeterminarea nu este eliminată de obicei pe temeuri *a priori*. Este greu de crezut totuși că ea are loc tot timpul, pentru că acest lucru ar implica o coincidență prea mare. Totuși, dacă noi considerăm că stările mentale și fizice sunt legate prin legi psiho-fizice – pretenție care poate fi apărată pe temeuri independente – supradeterminarea nu ar mai fi o coincidență: ar fi o lege naturală faptul că efectul este determinat atât de cauze mentale cât și fizice. Să presupunem că am o durere de cap ceea ce mă determină să mă duc la dulap și să iau o aspirină. Durerea și starea creierului sunt ambele cauze reale ale deplasării mele către dulap și, cum ele sunt legate de o lege psiho-fizică, nu este o coincidență faptul că ambele cauze au ca efect faptul că eu mă îndrept către dulap.

Să observăm că această perspectivă poate de asemenea înlătura consecința contra-intuitivă că m-aș fi îndreptat totuși către dulap chiar dacă nu aș fi avut o durere de cap. Pentru că nu ar fi adevărat că dacă nu aș fi avut acea durere m-aș fi dus totuși la dulap. Motivul este acela că, datorită faptului că durerea și starea creierului meu sunt legate de o lege psiho-fizică, cele mai apropiate lumi în care n-am avut durerea sunt acelea în care n-am avut nici acea stare a creierului<sup>42</sup>.

Acestea sunt două modalități non-fizicaliste de a rezolva problema cauzării mentale. Totuși, oricât de plauzibile sunt aceste opțiuni care nu se bucură de prea multă popularitate, răspunsul standard este respingerea lui (E), omogenitatea cauzării mentale și fizice. Astfel spus, dacă există cauze mentale ele nu sunt cauze „în același fel“ în care sunt cele fizice. Unii filosofi sunt mai expliciti decât alții în privința respingerii lui (E). Cei mai

expliciți sunt anumiți teoreticieni fizicaliști care susțin eficiența conținutului intențional. Dretske, de pildă, distinge în mod explicit între modul în care stările intenționale sunt cauze „structurante“ ale comportamentului și modul în care stările fizice sunt cauze „declanșatoare“. Iar Frank Jackson și Philip Pettit disting între tipurile de explicații pe care le oferim fenomenelor intenționale – „explicații program“ – și explicațiile cauzale obișnuite<sup>43</sup>. Dar, cum eficiența conținutului nu constituie tema acesui articol, nu am să discut aceste teorii.

O explicație fizicalistă influentă a cauzării mentale este teoria lui Kim conform căreia cauzarea mentală este „cauzare supervenientă“. În general,  $X$  îl cauzează în mod supervenient pe  $Y$  dacă și numai dacă  $X$  supervine pe  $X^*$  și  $Y$  supervine pe  $Y^*$  și  $X^*$  îl cauzează pe  $Y^*$ . În cazul mental, „un eveniment mental  $M$  cauzează un eveniment fizic  $P$ ... pentru că  $M$  supervine pe un eveniment fizic  $P^*$  și\* îl cauzează pe  $P$ “<sup>44</sup>. Oricare ar fi meritele acestei idei este clar că dacă completitudinea și respingerea supradeterminării sunt păstrate, atunci cauzarea supervenientă nu poate fi același lucru cu cauzarea fizică subsidiară. Din acest motiv poziția lui Kim implică o respingere a omogenității<sup>45</sup>.

O altă încercare de a reconcilia presuposițiile (A) – (E) este teza susținută de Block și Papineau că proprietățile mentale sunt proprietăți de al doilea grad: o proprietate mentală  $M$  este proprietatea de a avea o anumită proprietate  $P$ , a cărei instanțiere are anumite cauze și efecte<sup>46</sup>. Dacă se insistă, în pofida acuzației de epifenomenalism, că  $M$  este totuși „responsabilă“ de aceste efecte, atunci trebuie să întrebăm: în ce sens? Dacă supradeterminarea și completitudinea fizicii sunt respinse, nu poate fi același sens în care  $P$  este responsabil pentru efectele în discuție. Deci, omogenitatea este respinsă.

Aceași schemă generală se aplică și ingenioasei sugestii recente a lui Stephen Yablo că proprietățile mentale ale căror realizatori fizici sunt determinate sunt determinabile (în sensul lui W. E. Johnson). De exemplu, așa cum a fi colorat este un determinabil care are multe determinate (verde, roșu etc.), a simți o durere este un determinabil care are multe determinate – tipurile de proprietăți ale creierului care sunt realizatorii durerii. Soluția problemei cauzării mentale trimite deci la ideea că așa

cum nu există supradeterminare a efectelor unui obiect ca fiind verde, prin aceea că obiectele verzi sunt de asemenea obiecte colorate, nu există nici un risc de supradeterminare a efectelor proprietăților fizice de către proprietățile mentale pe care le realizează. Determinatele și determinabilele lor nu sunt „rivali cauzali”<sup>47</sup>.

Yablo respinge ideea că faptele mentale și fizice sunt identice. Dar, de vreme ce el consideră că teza lui este compatibilă cu completitudinea fizicii – „tot ceea ce se întâmplă este o consecință cauzală strictă a antecedentilor săi fizici”<sup>48</sup> – el trebuie să respingă și omogenitatea. Și face acest lucru distingând între „suficiența cauzală” și „relevanța cauzală”. O proprietate mentală poate fi cauzal relevantă pentru efectele instanțelor sale, dar numai proprietățile fizice sunt cauzal suficiente.

Uneori se obiectează că aceste răspunsuri variate la problema cauzării mentale sunt într-adevăr doar forme, ușor mascate, de epifenomenalism. Este însă greu de judecat această problemă fără a intra în detaliile teoriilor cauzării pe care aceste răspunsuri le presupun. Mai mult, fizicaliștii ortodocși înclină să răspundă că (de exemplu) cauzele superveniente sunt, totuși, cauze „într-un anume sens” și că numai dacă avem o concepție primitivă despre cauzare putem crede că, cauzele mentale trebuie să fie cauze în „același sens” ca cele fizice<sup>49</sup>. Din aceste motive, există tendința ca dezbaterea să degenereze într-un schimb nefertil de intuiții referitoare la ce înseamnă exact un „epifenomen”. Eu vreau să încerc o abordare diferită și să cercetez efectul pe care îl are respingerea omogenității asupra motivației fizicalismului ortodox.

Fizicalismul ortodox, așa cum văzut, dă naștere aceleiași probleme pentru a cărei rezolvare a fost introdus la început fizicalismul. Pentru a rezolva această problemă în termenii fizicalismului ortodox, manevra tipică este aceea de a respinge omogenitatea cauzării mentale și fizice. Dar, problema legată de respingerea omogenității este aceea că este imposibil acum chiar și numai să enunțăm motivația originală a fizicalismului: conflictul dintre cauzarea mentală și de completitudinea fizicii. Deci, nu exista nici un motiv clar pentru a spune că aceste fenomene mentale sunt „constituite de” sau „realizate de” fenomene fizice. Fizicalismul și-a pierdut din vedere motivația.

De fapt acest aspect este implicit în articolul lui Lewis, din 1966; numai atâta vreme cât stările mentale au efecte în *exact același sens în care stările fizice au efecte*, avem nevoie să le privim ca stări fizice. Așa cum am văzut, fizicalismul ortodox respinge faptul că stările mentale au efecte „în exact același sens” ca stările fizice. Pentru că el spune că există fenomene mentale care sunt cauze într-un mod propriu; și există fenomene fizice care sunt cauze în alt fel. Indiferent dacă aceste distincții între diferite tipuri de cauzare pot fi susținute sau nu, în ultimă instanță, pare clar că lucrurile nu se schimbă prea mult dacă spunem că aceste fenomene mentale sunt „în cele din urmă fizice”. Și pare o decizie terminologică lipsită de conținut să numim poziția care rezultă de aici „fizicalism” – cu excepția poate a consemnării diferențelor față de Descartes. Dar dacă asta este tot ce înseamnă fizicalismul, atunci nu mai este nevoie de formularea unei versiuni precise a doctrinei.

## X

**Concluzie.** Care este deci lecția pe care ne-o oferă disputa recentă asupra cauzării mentale? Nu neg faptul că cineva ar putea aduce o explicație filosofică a relației dintre minte și corp – pe linia constituire / realizare superveniență – pe care să o numească „fizicalism”. Părerea mea este, totuși, aceea că aceste varietăți de fizicalism nu au o motivație filosofică clară. Lecția care decurge din disputa cauzării mentale este aceea că nu există o poziție fizicalistă binemotivată care să nu fie o teorie a identității.

Această concluzie rezolvă dificultatea noastră inițială. Dar rămâne, totuși, o altă dificultate. În ultima vreme, filosofia minții a fost dominată de încercări de a descrie și apăra versiuni adecvate ale „fizicalismului non-reducționist”. Aceste încercări sunt adesea răspunsuri la problema cauzării mentale, dar ele sunt propuse ca prezentând un interes filosofic independent. Ceea ce reiese clar din această dezbatere este că relația de cauzare mentală constituie o problemă numai pentru cei care resping teoria identității.



Cu toate acestea, dat fiind răspunsul standard la această problemă, nedumerește faptul că fizicaliștii consideră că întrebarea „care este versiunea adecvată a fizicalismului?” are încă nevoie de un răspuns. Că aceasta este o întrebare semnificativă, fizicaliștii trebuie să explice, totuși, de ce<sup>50</sup>.

## NOTE

<sup>1</sup> Vezi Stephen Yablo, „Mental Causation“, în *Philosophical Review*, 101, 1992, 3; Philip Pettit, „A Definition of Physicalism“, în *Analysis*, 53, 1993; și Jeffry Poland, *Physicalism: the Philosophical Foundations*, (Oxford, Clarendon Press, 1994).

<sup>2</sup> Vezi (e.g. Terence Horgan, „From Supervenience to Superdupervenience: Meeting the Demands of a Material World“, în *Mind*, 102, 1993; Poland, *Physicalism*, p. 105; David Charles, „Supervenience, Composition and Reduction“, în D. Charles și K. Lennon, *Reduction, Explanation and Realism* (Oxford, Clarendon Press, 1992).

<sup>3</sup> Vezi J. Kim, „The Myth of Non-Reductive Materialism“, în *Supervenience and Mind*, (Cambridge, Cambridge University Press, 1993); Brian Loar, „Elimination versus Non-Reductive Physicalism“, în Charles și Lennon, editori, *Reduction, Explanation and Realism*.

<sup>4</sup> Vezi U.T. Place, „Is Consciousness a Brain Process?“, (p. 42) și J.J.C. Smart, „Sensations and Brain Processes“, (p. 54), în C.V. Borst (ed.), *The Mind-Brain Identity Theory*, (Londra, Macmillan, 1970).

<sup>5</sup> *A Materialist Theory of Mind*, (London, Routledge, 1993), p. 82. Argumentul lui Armstrong este la pp. 82–90.

<sup>6</sup> „An Argument for the Identity Theory“, în *Philosophical Papers Volume I*, (Oxford, Oxford University Press, 1983), p. 105.

<sup>7</sup> Vezi James Hopkins, „Mental States, Natural Kinds and Psychophysical Laws“, în *Proceedings of the Aristotelian Society Supplementary Volume*, 1978; Christopher Peacocke, *Holistic Explanation* (Oxford, Clarendon Press, 1979) și David Papineau, „Why Supervenience?“, în *Analysis*, 50, 1990.

<sup>8</sup> Kim, J. „The Myth of Non-Reductive Materialism“, p. 281; Peacocke, *Holistic Explanation*, p. 135.

<sup>9</sup> Stephen Schiffer, *Remnants of Meaning* (Cambridge Mass., MIT Press, 1987), p. 148.

<sup>10</sup> Donald Davidson, „Mental Events“, în *Essays on Actions and Events* (Oxford, Oxford University Press, 1980).

<sup>11</sup> David Papineau, *Philosophical Naturalism*, (Oxford, Blackwell, 1993), p. 16. Vezi și Horgan, „From Supervenience to Superdupervenience“, p. 573.

<sup>11</sup> Conf. Hopkins: „Nici... problema alegerii între dualism și fizicalism nu este lipsită de conținut empiric. Pe scurt, problema este dacă evenimentele mentale și fizice interacționează și, dacă acceptăm acest lucru, dacă partea relevantă a lumii fizice este cauzal completă.“, „Mental States, Natural Kinds and Psychophysical Laws“, p. 225.

<sup>12</sup> John Searle, *Intentionality*, (Cambridge, Cambridge University Press, 1983), p. 265.

<sup>13</sup> Versiuni ale teoriei constituirii sunt acceptate și susținute de mulți dintre autorii care au colaborat la două antologii recente, *Mental Causation*, editori J. Heil și A. Mele, (Oxford, Oxford University Press, 1991) și *Reduction, Explanation and Realism*, editori D. Charles și K. Lennon. Pentru o expunere excelentă a problemei, vezi Paul Snowdon, „On formulating materialism and dualism“, în *Cause, Mind and Reality*, ed. J. Heil, (Dordrecht, Kluwer, 1989).

<sup>14</sup> Vezi D. Charles, „Supervenience, Composition and Reduction“; Pettit, „A Definition of Physicalism“; Horgan, „From Supervenience to Superdupervenience“, p. 566.

<sup>15</sup> Vezi Poland, *Physicalism*, capitolul 3.

<sup>16</sup> Pentru concepția lui Kim vezi „Causation, Nomic Subsumption and the Concept of Event“ și „Events as Property Exemplifications“, în *Supervenience and Mind*. Eu urmez perspectiva lui Davidson, din „The Individuation of Events“, în *Essays on Actions and Events*, care îl urmează pe Ramsey, „Facts and Propositions“, în *Philosophical Papers*, (Cambridge, Cambridge University Press, 1990).

<sup>17</sup> Pentru o bună prezentare a acestui punct, vezi Papineau, *Philosophical Naturalism*, p. 24.

<sup>18</sup> Vezi Cynthia și Graham Macdonald, „Mental Causes and the Explanation of Behaviour“, în *Philosophical Quarterly*, 1986; „Mental Causation and Non-Reductive Monism“, în *Analysis*, 51, 1991; Cynthia Macdonald, *Mind-Body Identity Theories*, (London, Routledge, 1989), capitolele 4 și 5.

<sup>19</sup> Vezi „Mental Causation and Non-Reductive Monism“, pp. 27–28. Pentru tropi, vezi Keith Campbell, „The Metaphysics of Abstract Particulars“, în *Midwest Studies in Philosophy VI: The Foundations of Analytic Philosophy*, (Minneapolis, University of Minnesota Press, 1981).

<sup>20</sup> Vezi Kim, „Causation, Nomic Subsumption and the Concept of Event“.

<sup>21</sup> „Mental Causation and Non-Reductive Monism“, p. 28. Pentru detalii referitoare la metafizica acestei perspective, vezi Cynthia Macdonald, „Constitutive Properties, Essences and Events“, *Philosophia*, 16, 1986. Teoria merită să fie discutată pe larg.

<sup>22</sup> Vezi John Heil, *The Nature of True Minds*, (Cambridge, Cambridge University Press, 1992), pp. 135–139, care susține o teorie foarte asemănătoare cu cea a soților Macdonalds și cu versiunea ortodoxă a fizicalismului. El o numește „realizare“.

<sup>23</sup> „A Definition of Physicalism“ în *Analysis*, 53, 1993, p. 215.

<sup>24</sup> Vezi Ernest Sosa, „Mind-Body Interaction and Supervenient Causation“, în *Midwest Studies in Philosophy* 9, (1984), p. 278; Brian McLaughlin, „Type Epiphenomenalism, Type Dualism and the Causal Priority of the Physical“, în *Philosophical Perspectives*, 3; *Philosophy of Mind and Action Theory*, editor James E. Tomberlin (Atascadero, Ridgeview, 1989), și Yablo, „Mental

Causation“, pp. 248–250. Așa cum observa Yablo, distincția îi aparține lui C. D. Broad, *The Mind and its Place in Nature*, (London, Routledge & Kegan Paul, 1925), p. 473.

<sup>26</sup> Pentru o expunere clară a problemei eficienței conținutului, vezi Ned Block, „Can the Mind Change the World?“, în *Meaning and Method*, editor George Boolos, (Cambridge, Cambridge University Press, 1990). Unii autori care discută problema eficienței conținutului abordează de asemenea și problema cauzării mentale, din a cărei perspectivă pot fi interpretați cel mai bine: vezi, de exemplu, Gabriel Segal și Elliot Sober, „The Causal Efficacy of Content“, în *Philosophical Studies*, 63, 1991, și Michael Tye, *The Imagery Debate*, (Cambridge Mass., MIT Press, 1992), capitolul 8.

<sup>27</sup> Pentru mai multe detalii despre criteriile contrafactice ale „relevanței cauzale“, vezi Ernest Le Pore și Barry Loewer, „Mind Matters“, în *Journal of Philosophy*, 84, 1987; Jerry Fodor, „Making Mind Matter More“, în *A Theory of Content and Other Essays*, (Cambridge, Mas., MIT Press, 1990); Gabriel Segal și Elliot Sober, „The Causal Efficacy of Content“, și Brian Leiter și Alexander Miller, „Mind Doesn't Matter Yet“, în *Australian Journal of Philosophy*, 72, 1994.

<sup>28</sup> Vezi Frank Jackson, „Essentialism, Mental Properties and Causation“, în *Proceedings of the Aristotelian Society*, XCV, 1995. Eu am altă părere decât Jackson în privința structurii problemei cauzării mentale. Dar sunt de acord cu el în privința relevanței proprietăților în cadrul cauzării.

<sup>29</sup> Vezi Ted Honderich, „The Argument for Anomalous Monism“, în *Analysis*, 42, 1982; și, de asemenea, Ernest Sosa, „Mind-Body Interaction and Supervenient Causation“.

<sup>30</sup> „Thinking Causes“, în Heil și Mele, (editori), *Mental Causation*, p. 13.

<sup>31</sup> „Thinking Causes“, p. 12.

<sup>32</sup> Îi sunt recunoscător lui Alex Oliver care mi-a atras atenția asupra problemei discutate aici.

<sup>33</sup> Vezi, de exemplu, Horgan, „From Supervenience to Superdupervenience“, p. 563.

<sup>34</sup> Vezi Jackson, „Essentialism, Mental Properties and Causation“.

<sup>35</sup> Pentru critica abordării acestei probleme din perspectiva instanței-proprietate, vezi Yablo, „Mental Causation“, p. 259, n32, și E.J. Lowe, „The Causal Autonomy of the Mental“, în *Mind*, 102, 1993, p. 631.

<sup>36</sup> Jerry Fodor, *Psychosemantics*, (Cambridge, Mass., MIT Press, 1987), p. 30. Schiffer critică și el acest aspect în *Remnants of Meaning*, p. 154.

<sup>37</sup> Vezi e.g., David Papineau, „The Reason Why: Response to Crane“, în *Analysis*, 52, 1991.

<sup>38</sup> Tyler Burge, „Mind-Body Causation and Explanatory Practice“, în Heil & Mele (editori), *Mental Causation*, p. 115. Vezi, de asemenea, E.J. Lowe, „The Causal Autonomy of the Mental“, p. 637.

<sup>39</sup> Vezi Pettit, „A Definition of Physicalism“, p. 219; Horgan, „From Supervenience to Superdupervenience“, p. 572.

<sup>40</sup> Pentru o discuție excelentă, vezi Burge, „Mind-Body Causation and Explanatory Practice“, în special pp. 111–117..

<sup>41</sup> Vezi, de exemplu, William Seager, *The Metaphysics of Consciousness*, (Londra, Routledge, 1991), p. 31–32.

<sup>42</sup> Ideea îi aparține lui Hugh Mellor: vezi D.H. Mellor, *The Facts of Causation*, (Londra, Routledge, 1995), capitolul 8.

<sup>43</sup> Vezi Fred I. Dretske, *Explaining Behaviour*, (Cambridge Mass., MIT Press, 1988); Franck Jackson și Philip Pettit, „Functionalism and Broad Content“, *Mind*, 97, 1988.

<sup>44</sup> Vezi Kim, „Epiphenomenal and Supervenient Causation“, în *Supervenience and Mind*, p. 106. Superveniența în discuție este superveniența „tare“. Vezi, de asemenea, Segal și Sober, „The Causal Efficacy of Content“, și Tye, *The Imagery Debate*, capitolul 8, în special pp. 147–8.

<sup>45</sup> De fapt, Kim pare să oscileze între respingerea lui (B) și respingerea lui (E)- și asta pentru că el consideră cauzarea supervenientă „cauzare epifenomenalistă“: vezi, „Epiphenomenal and Supervenient Causation“.

<sup>46</sup> „Can the Mind Change the World?“, pp. 155–166; în *Philosophical Naturalism*, p. 25.

<sup>47</sup> Yablo, „Mental Causation“, p. 259. Concepția lui Yablo merita o discuție mult mai detaliată decât îmi permite articolul de față.

<sup>48</sup> Yablo, „Mental Causation“, p. 279.

<sup>49</sup> Vezi Dennett, „Real Patterns“, în *Journal of Philosophy*, 88, 1991, p. 43.

<sup>50</sup> Le sunt foarte recunoscător lui Doug Ehring, Peter Goldie, Keith Hossack, Hugh Mellor, Lucy O'Brien, Alex Oliver, David Papineau, Sarah Patterson, Gabriel Segal, Barry Smith, Scott Sturgeon, Bernhard Weiss, Nick Zangwill și în special lui Mike Martin, pentru discuțiile în care au contribuit la scrierea acestui articol. Versiuni mai vechi ale unor părți din acest material au fost prezentate în cadrul unor conferințe publice, la universitățile din Birmingham, Cambridge, Leeds, Liverpool, Manchester și Wales (Lampeter). Le sunt recunoscător participanților la acele discuții pentru reacțiile și comentariile lor.

Traducere Laurențiu STAICU

# NATURALISM ȘI PROBLEMA INTENȚIONALITĂȚII \*

*John J. HALDANE*

(Universitatea St. Andrews, Anglia)

## ÎN MEMORIA LUI IAN McFETRIDGE 1948–1958

Eseul<sup>1</sup> se ocupă de problema dacă stările cognitive pot fi explicate în termeni naturaliști (adică fizicaliști). Argumentul prezentat în favoarea unui răspuns negativ derivă din ideea că stările cognitive implică moduri de prezentare ale căror condiții de identitate și individualizare sunt în mod cert intenționale și intensionale și, în consecință, nu pot fi explicate în termenii stărilor fizico-cauzale. În legătură cu tema este examinată și respinsă încercarea recentă a lui Robert Stalnaker de a realiza o teorie naturalistă a intenționalității. În concluzie se sugerează că soluția mai radicală la problema modalității de a corela fenomenele cognitive și cele materiale poate fi găsită în naturalismul pluralist al lui Aquinas (și probabil al lui Wittgenstein).

### I

Naturalismul filozofic este o abordare veche a temei descrierii și explicării acelor fenomene foarte generale a căror înțelegere cade în sarcina reflecției a priori și nu (sau nu numai) a științei. Ca urmare, există tot atâtea naturalisme câte câmpuri de cercetare filosofice și este important să observăm că, dacă am putea fi

---

\* J. J. Haldane, „Naturalism and the Problem of Intentionality” în *Inquiry. An Interdisciplinary Journal of Philosophy* 32/3 sept. 1989.

satisfăcuți de o explicație naturalistă a unui șir de fenomene, cele morale să zicem, s-ar putea, în același timp, și pentru aceleași motive, să ajungem să preferăm o explicație non-naturalistă a elementelor în alt domeniu – mă gândesc la exemplul matematicii. Astfel, naturalismul, ca și realismul, nu este același lucru peste tot sau oriunde. Cele două abordări sunt și asemănătoare, prin aceea că ambele pot fi avansate ca generalități; și este posibil ca principalele motive care au dus la preferarea lor într-un domeniu anume să fie cele care se aplică global sau, mai puțin riguros spus, că succesul uneia sau alteia dintre abordări într-un câmp dat devine o bază pentru a o promova ca pe o posibilă poziție general explicativă. Un prim exemplu pentru naturalism ar fi concepția potrivit căreia singura abordare explicativă a cunoașterii ar fi tratarea ei ca un fenomen cauzal, care implică receptarea emisiilor răspândite de entități fizice într-un mediu material sau versiunea actuală a epistemologiei atomiste care implică și procesele de cunoaștere. Dacă această supoziție ar fi adevărată, atunci s-ar părea că naturalismul global se impune nu întrucât toate exemplele de non-naturalism s-ar dovedi false (putem să ne întrebăm dacă există un domeniu platonice al obiectelor matematice de exemplu), ci pentru că ar lipsi o motivație majoră pentru non-naturalism. Problema cunoașterii fiind rezolvată, ce altă problemă ar mai putea rămâne de rezolvat? Mediul cercetării filosofice este gândirea, indiferent care subiect particular al ei poate fi disponibil drept obiect al cunoașterii, actual sau posibil. Așa că, dacă naturalismul de tipul pe care l-am menționat constituie abordarea corectă a cogniției, atunci naturalismul global pare singura opțiune rezonabilă.

Calea mai puțin riguroasă a naturalismului de la particular la general este sugerată de metodele de investigație inductive, mai precis de economia ontologică. Distincția între natural și non-natural trebuie precizată, fiind foarte clar că unele lucruri aparțin primei categorii și foarte evident care sunt ele. Așa stând lucrurile, naturalismul apare drept abordarea corectă pentru a înțelege anumite fenomene recunoscute ca generale. Depărtarea de el implică expansiunea ontologică și eșecul prezumției uniformității. În acest sens, deci, non-naturalismul aduce vești rele susținând că povestea realității va include, în orice caz,

varietatea calitativă și posibilitatea ca ea să nu poată fi încheiată ca o singură poveste. Din nou naturalismul dintr-un domeniu, oferă motive pentru acceptarea lui generală, deși, în cazul de față soliditatea motivelor este mai puțin evidentă.

Până acum nu am spus exact ce consider că este naturalismul, deși observațiile mele sugerează o noțiune larg folosită, echivalentă cu fizicalismul: punctul de vedere că realitatea sau, oricum, domeniul particularilor concreți, al proprietăților și al relațiilor sunt similare cu lumea cercetată de științele fizice. Punctul de vedere comun astăzi filosofilor de limbă engleză reprezintă un fel de ortodoxie. Deoarece deși în ultimul timp s-a scris mult despre, și, împotriva reduccionismului, de exemplu, în etică și în filosofia minții, tezele atacate sunt de obicei cele conceptuale sau semantice, rareori exprimându-se îndoieli asupra faptului că la bază realitatea este fizică. Monismul anomalistic și alte versiuni ale fizicalismului simbolic sunt larg acceptate datorită faptului că reconciliază fenomenele psihologice cu o ontologie naturalistă, adică fizicalistă.

Înțelegerea actuală a naturalismului se cere deosebită de înțelegerile, asociate într-o oarecare măsură cu Aristotel și dezvoltate cu entuziasm de discipolii săi. Concepția aristotelică derivă din insatisfacția față de dualismul platonice ca abordare a câmpului gândirii și experienței, și ca atare înclină către monism, dar nu este fizicalistă. În perspectiva lui Aquinas, de exemplu, naturalismul se extinde, insistând asupra unității omului – ca o unică substanță psihofizică<sup>2</sup> și a lui Christos – ca o singură persoană theoanthropică. Naturalismul, în acest al doilea sens, este o abordare la care nu am nici o obiecție. Deși se confruntă cu multe probleme, sunt înclinat să optez pentru el, întrucât reprezintă o alternativă atractivă la excesele platonismului și la restricțiile fizicalismului.

După cum rezultă din ultima observație, eu mă opun naturalismului în modul în care este conceput curent. Orice filosof ar trebui să fie cel puțin suspicios în legătură cu el, deoarece, dacă naturalismul științific ar fi adevărat, ar implica și conceperea cunoașterii ca un fenomen cauzal de tipul menționat anterior și nu ar mai fi clar cum e posibilă filosofia ca o cercetare a priori. Naturalismul se află în spatele recentelor atacuri împotriva

apriorismului, avansate de filosofii de orientare radical materialistă. Dar naturalismul se poate la fel de ușor contrapune materialismului, tratându-l ca pe o infirmare a premisei privind cunoașterea drept fenomen natural în sens restrâns.

Nu vom încerca să stabilim că filosofia reprezintă un studiu a priori și că aprioricul nu poate fi considerat cauzal, în vreme ce naturalismul științific este obligat la o abordare cauzală a cogniției, ci ne vom restrânge de aici încolo la problema dacă stările cognitive, în genere, pot fi interpretate în termeni naturaliști și la respingerea unei încercări recente de a arăta că se poate.

## II

Așa cum folosesc eu termenul, o stare este „cognitivă” dacă intrinsec se dovedește plină de conținut, adică, dacă fiind într-o astfel de stare o existență este, ipso facto, conștientă de unele elemente reprezentationale. Nu intenționăm o analiză reductivă ci doar introducerea unor noțiuni înrudite. Îndoielile asupra inteligibilității ideii de voință cognitivă nu vor fi anulate de menționarea „conținuturilor” sau a „reprezentărilor”, dar acestea din urmă indică cel puțin felul lucrului care se află sub denumirea de *stare cognitivă*. Astfel, gândurile, cerințele, percepțiile, imaginile și experiențele non-relaționale de exemplu, durerile, impresiile de culoare etc. sunt toate cognitive, potrivit sensului acestui termen.

Ceea ce nu intră aici, sunt pur și simplu stările care constau din sau sunt consecințe cauzale ale reacțiilor fizice sistematice la mediu, ceea ce unii filosofi preferă să numească stări informaționale. În afară de materialismul radical, care nu acordă nici un loc intenționalității, în general este recunoscut faptul că există o diferență între stările informaționale, ca de exemplu, temperatura fluctuantă a camerei (funcție a temperaturii de afară și a căldurii corpurilor din interior) și stările cognitive, ca de exemplu, senzația de căldură sau credința că temperatura are variații sistematice. Ceea ce se dispută totuși, se referă la natura



diferenței. Unii autori o privesc ca pe o cheștiune de grad de complexitate. Eu vreau să argumentez că este o diferență de calitate. Problema cognitivului nu poate fi soluționată la nivel informațional sau, ceea ce este același lucru, concepută ca simplă sensibilitate la mediu. Trecând la perspectiva istorică, pretenția de diferență radicală poate fi pusă în termenii dezvoltării discontinue negând cognitivului (cu proprietățile care îl însoțesc) posibilitatea de a fi reprezentat ca derivând prin procese naturale din materia non-conștientă (și non-intențională) pre-existentă, chiar dacă ea conține stări informaționale.

Problema se leagă direct de aspectele fenomenologice ale evenimentelor mentale. Aici am să mă ocup de intenționalitate mai curând decât de „caracterul calitativ” al stărilor și proceselor psihologice care în literatura recentă mai mult decât conținutul conceptual este privit drept o problemă insurmontabilă pentru teoriile fiziciste ale minții. Evit subiectul „qualia” (*calități*) în primul rând, deoarece se dovedește irelevant pentru explicarea mentalității subconștiente, în timp ce intenționalitatea caracterizează orice activitate cognitivă, și în al doilea rând, deoarece nu sunt așa de sigur ca alții că am o idee clară despre ceea ce ar putea fi „qualia”. Mai mult, chiar dacă am o înțelegere reflectivă a naturii aspectelor senzoriale ale experienței, nu îmi este încă clar dacă ele trebuie să reziste în sine la încadrarea într-o ontologie materialistă. Dimpotrivă, sunt chiar convinși că puterile reprezentationale ale minții sunt inexplicabile în termenii de care dispune materialismul.

### III

Există desigur o legătură între cogniție și conștientă, dar ea nu a fost suficient explorată. Pe de o parte, Descartes a sugerat eronat că toate stările cognitive sunt ipso facto conștiente (Nimic nu există în minte, în măsura în care este un lucru gândibil, care să nu fie conștient<sup>3</sup>); iar la cealaltă extremă, Fodor, referindu-se

la intențional, concepe drept accidental faptul că în noi puterea reprezentării mentale este acompaniată de conștiință („s-ar putea ca să scoți din posibilitatea introspecției, posibilitatea cogniției conștiente... / dar / nu avem o explicație clară etiologică... de ce acele lucrări care sunt conștiente, sunt conștiente și celelalte nu sunt“). Nici una din aceste atitudini nu se dovedește satisfăcătoare pentru că nu toate stările intenționale sunt și conștiente, dar nici nu pare corect să privim legătura dintre gândire și conștiință ca fiind doar întâmplătoare. Pe de o parte, dacă ocolim ineismul avem nevoie de o explicație a genezei conținutului gândirii în individ, adică de o explicație a formării conceptului și este posibil ca ea să implice ideile atenției selective și a instrucției sociale. Pe de alta, o mare parte din conținutul gândirii se dovedește direct auto-referențial și presupune un subiect cognitiv înzestrat cu putere de reflecție. Așa că, nu este deloc clar ce ar putea rămâne de sine stătător, dacă ar rămâne ceva, după ce conștiința ar fi îndepărtată de pe scena gândirii.

Conștiința pare a fi legată de cogniție în două feluri. Primul corespunde cazurilor deja menționate și reprezintă formarea anumitor clase de conținuturi, inclusiv cele evident experimentale și reflectorii. Al doilea apare drept o formă mai profundă de implicare. Mai înainte am comparat stările informaționale cu cele cognitive și am propus să fie considerate calitativ diferite. Dar aceasta nu exclude posibilitatea ca stările cognitive să absoarbă pe primele în calitate de constituenți parțiali. Percepția s-ar părea că se prezintă în mod clar ca un astfel de caz. În dimensiunea sa vizuală, ochii, nervii optici și alte aspecte neuroanatomice relevante răspund în modalități sistematice la schimbările mediului și, astfel, se poate spune că ele codifică informația despre mediu. Dar aceasta nu reprezintă încă percepția. Stărilor astfel produse le lipsește conținutul conceptual, caracterul fenomenologic și conștiința persoanei; ele reprezintă sensibilitatea productivă ambientală, un fenomen în întregime cauzal, care poate fi caracterizat din punct de vedere non-cognitiv. În ce constă deci diferența dintre această stare și percepție sau mai bine spus: cine efectuează trecerea de la sensibilitatea organică la activitatea perceptuală? Se autoimpune ca răspuns: conștiința reflexivă. În virtutea prezenței ei, acel lucru care altfel ar fi doar *cauza* unei

stări informaționale, este perceput ca *obiect* al experienței. Motivul cheie al includerii reflexivității în această caracterizare a conștiinței este faptul că ea aduce recunoașterea conștiinței de sine și astfel a posibilității credinței, valorizării, revizuirii și a altor trăsături asociate cu gândirea rațională<sup>5</sup>.

Observați și faptul că, odată ce se introduce conștiința, încercările de a caracteriza stările nou instaurate, trebuie să specifice atât trăsăturile relevante ale mediului extern de care este legat subiectul cât și modul în care se prezintă respectivele trăsături celui care le percepe. Adică, astfel de atribuiri de conținut trebuie să fie sensibile și la modurile cognitive de prezentare prin care este experimentată lumea. Implicația acestei cerințe este că, în timp ce în împrejurările primare de producere a stărilor informaționale, relația dintre mediu și determinările senzațiilor relevante a fost cauzală și extensională, relația cognitivă, ca relație de referință mediată prin conținutul conceptual (și fenomenologic), este în mod clar ne-extensională<sup>6</sup>.

#### IV

Am remarcat anterior că ideea ireductibilității stărilor cognitive la cele informaționale, poate apărea deghizată în perspectivă istorică sub forma imposibilității explicării stărilor cognitive ca fiind formate prin procese naturale din cele informaționale. Un argument în acest scop poate fi prezentat pe baza observației recente cu privire la extensionalitatea descripțiilor relațiilor cognitive. Există o dispută bine cunoscută care își are originile în scrierile lui *Chisholm* și *Sellars* cu privire la posibilitatea definirii fenomenelor mentale prin referință la non-extensionalitatea pretinsă a propozițiilor psihologice și la problema dacă fenomenele mentale fiind intensionale, ar putea fi reduse la altele cu care sunt echivalente din toate punctele importante de vedere, dar care sunt extensionale în mod neproblematic<sup>7</sup>. Argumentul meu se corelează acestei perspective dar conectează intensionalitatea

atribuirilor de conținut cu problema dacă acel lucru de care depinde conținutul, adică conștiința, s-ar fi putut dezvolta din materia fără spirit. O astfel de posibilitate constituie o condiție necesară, deși nu suficientă pentru succesul naturalismului științific.

Există mai multe puncte de vedere filosofice despre ontologia și despre logica corectă a teoretizării științifice, dar cea care a fost și rămâne cea mai influentă este bine cunoscuta poziție derivată din pozitivismul logic, care combină teza extensionalității a lui Carnap<sup>8</sup> cu modelul de explicație „lege – de – acoperire” – a lui Hempel (mai ales în varianta ei deductivă – problema legilor statistice nefiind relevantă pentru problema prezentă)<sup>9</sup>. Naturalismul științific astfel conceput caută să dezvolte o descriere completă a realității, suficientă să explice apariția tuturor fenomenelor. Idealizând, pornind de la împrejurările practic posibile sau reale, el susține că, date fiind, ca premise, o judecată despre condițiile inițiale ale universului și setul de legi care determină ce se poate întâmpla în interiorul lui, se poate obține, deductiv, descrierea oricărui fenomen care apare ulterior și prin aceasta explicarea lui. După cum s-a observat, teoria presupune un limbaj extensional. Totuși, discuția anterioară a clarificat faptul că o parte din descrierea psihologică a persoanelor trebuie să țină cont de modalitățile experienței lor: de conceptele prin care ele gândesc lucrurile și de caracterul fenomenal al percepțiilor lor. Asemenea descripții sunt inevitabil intensionale. Nici o judecată în întregime extensională nu antrenează una intensională și deci, nu poate exista nici explicația deductivă a apariției stărilor intenționale sau a evoluției a ceea ce presupun ele, adică a conștiinței reflexive. Pe scurt, deși se presupune că psihologicul reprezintă o rezultantă cauzală a condițiilor fizice anterioare, naturalismul ortodox este *logic* incapabil să explice apariția fenomenelor intenționale. Se poate obiecta că acest raționament presupune că intenționalitatea este o notă distinctivă a mentalului, când de fapt, nu este nici necesară și nici suficientă pentru el. În ce privește prima condiție, eu nu cred că toate fenomenele psihologice sunt intenționale (senzațiile nu au astfel de trăsături), și apoi, descrierile intenționale nu interzic interferența existențială și substituția *salva verita*<sup>10</sup>. Desigur, dacă atribuirea de conținut în genere vrea să fie fidelă conținutului stărilor psihologice ale subiectului – așa cum trebuie

să fie dacă urmează să ne informeze despre *ele* și astfel, să ne permită să raționalizăm acțiunile lui – atunci intenționalitatea este de neînlăturat.

La întrebarea dacă intenționalitatea reprezintă o condiție suficientă pentru contextele mentale, este mai greu de răspuns. Un răspuns la argumentul anterior ar putea fi să admitem imposibilitatea psihologiei științifice de tip pozitivist, dar să propunem versiuni mai plauzibile ale naturalismului care să fie limitate la non-extensionalitate, independent de orice încercare de a acomoda fenomenele psihologice. Conchidem deci că, la urma urmei, nu există nici o prăpastie logică între propozițiile care descriu realitatea eliminând conștiința și acelea care o descriu incluzând-o. Modalitatea și judecățile exprimând legile naturale și relațiile cauzale crează, în genere, contexte intensionale fără să implice intenționalitatea – cam așa se poate formula ideea care sprijină propunerea prezentă. (Desigur, judecățile de acest fel, sunt rostite de subiecți psihologici, dar *conținutul* lor nu este necesar să se ocupe sau să presupună mentalul).

Răspunsul introduce mai multe probleme decât pot fi tratate pe larg aici. Desigur, se poate arăta că eșuează încercarea de a submina argumentul împotriva psihologiei științifice. Pentru a vedea de ce, vom aborda pe scurt natura și scopul extensionalismului și vom examina cauza inadecvării lui în contexte cu atribuire de conținut intențional. Pentru ca o teorie să fie extensională se cere ca semantica ei să fie dată în termenii noțiunilor corelative de referință și obiecte de referință. Deși, de obicei, poți fi extensionalist fără a fi fizicalist, nu e așa de clar că poți susține în mod coerent fizicalismul fără o constrângere subînțeleasă la extensionalism – posibilă, cel puțin, în principiu. Oricum, extensionalismul este specific fizicalismului, deoarece acesta oferă perspectiva explicării deopotrivă a faptelor semantice și a explicației însăși, sub forma comportamentului de teoretizare lingvistică, recurgând numai la relațiile fizice între obiectele fizice.

Totuși, un extensionalism pur ca cel de mai sus se confruntă cu o problemă cunoscută când explică semantica termenilor generali. Tratarea din perspectiva lui a predicatelor ca desemnând colecții de obiecte sau a indivizilor ca membri ai acestor colecții este inadecvată cel puțin din două motive. Mai întâi,

predicatele coextensive pot să difere ca valoare semantică deoarece pot individualiza obiecte în cadrul aceleiași colecții, dar prin intermediul unor trăsături diferite. Se știe bine că toate ființele care au inimi, au și rinichi, dar a avea o inimă nu este același lucru cu a avea un rinichi și, în consecință, există o diferență similară și între predicatele corespunzătoare. În al doilea rând și, ca o consecință a primului punct, extensionalismul pur interzice descrițiile și explicațiile relațiilor cauzale care implică propri-etăți. Așa cum a avea o inimă diferă de / proprietatea de / a avea un rinichi, chiar dacă coincide cu ea, a produce sau a suferi un anumit efect în virtutea faptului că avem o inimă diferă de a acționa sau a suferi o acțiune, prin intermediul / proprietății de / a avea rinichi. Recunoașterea acestor adevăruri cere abandonarea extensionalismului pur în favoarea unei semantici care permite distincții mai subtile între predicate. Totuși, dacă este adevărat că o teorie științifică adecvată trebuie concepută altfel de cum o face pozitivismul ortodox, adică independent de felul în care privește el fenomenele psihologice, acest lucru nu servește la umplerea prăpastiei la care ne-am referit anterior. Deoarece singurul lucru pe care îl pretinde teoria termenilor generali de la un extensionalist este ca să-și revizuiască principiile de identitate și individualizare pentru predicate, astfel încât să permită distincția între proprietăți diferite comaterializate. Se susține deseori că acest lucru necesită introducerea unor entități care să aibă legătură cu mintea, ca semnificații, sensuri sau concepte, dar de fapt nu este necesar. Presupunând desigur că termenii implicați nu sunt ei înșiși descriții ale fenomenelor intenționale, un extensionalism modificat ar fi încă posibil, prin individualizarea predicatelor extra-mentale de la extremitatea relației lume-limbaj. Astfel, se poate susține, de exemplu, că două elemente lingvistice sunt actualizări ale aceluiași predicat dacă și numai dacă indică aceeași proprietate sau că sunt echivalente semantic dacă și numai dacă sunt coextensive în mod necesar. Ambele propuneri au nevoie de suplimentare - în primul caz printr-un criteriu ne-lingvistic - al identității de proprietate", în al doilea caz, de o interpretare a naturii necesității implicate. Totuși, pentru scopul prezent, este destul să observăm că abandonarea unui extensionalism pur, care interpretează predicatele în mod simplu, prin referința la colecții de particulari nu trebuie să ajungă la adoptarea unei semantici bazate pe intenții, deja astfel echipată pentru a găzdui fenomenele psihologice.

Prin contrast, problema intenționalității nu poate fi rezolvată decât admitând existența unor entități de un tip absolut diferit decât cele recunoscute de adepții unui extensionalism pur sau modificat, care sunt *modurile de reprezentare*. Fenomenele intenționale sunt în mod esențial, pline de conținut din punct de vedere conceptual. Ele implică o caracterizare a obiectului prin ceea ce intenționează. A gândi un *a* este întotdeauna echivalent cu a-l gândi într-un mod sau altul – ca un *F*, sau ca fiind *g* deci ca actualizare a unui concept care îl tipizează sau îl caracterizează. Tocmai acest aspect conceptual al referinței mentale (și lingvistice) este cel care generează intensionalitatea specială, ireductibilă, a contextelor intenționale, și care prezintă o problemă insurmontabilă pentru psihologia științifică. Chiar în forma ei îmbunătățită, psihologia caută să explice realitatea prin intermediul unei teorii, a cărei semantică poate fi furnizată fără a recurge la alte elemente în afară de obiecte, proprietăți și relații naturale. Totuși, în interpretarea semnificației și adevărului atribuirii de conținut, este necesar să recunoaștem existența unei alte clase de entități: conceptele. Desigur, conceptele sunt proprietăți (mentale) dar ele diferă de atributele non-psihologice în două moduri importante. Mai întâi, ele sunt mai subtil individualizate; apoi ele sunt în mod esențial (re) prezentaționale. În ce privește prima diferență, trebuie să fie foarte clar că este nu numai imposibil să identifiți sau să deosebiți conceptele prin extensiunile lor reale, dar ele rămân nedeterminate chiar când sunt introduse și proprietățile naturale. Două expresii pot fi coextensive în mod necesar sau să vizeze aceeași proprietate fiind, totuși, distincte. Nici o încercare de a adăuga alte restricții din partea lumii exterioare minții, adică din domeniul non-intențional nu ar putea să servească la individualizarea conceptelor, deoarece există un număr nelimitat de posibilități de reprezentare a aceluiași lucru. Așa încât nici o acumulare de elemente non-intenționale în nici un mod neintențional de combi-nare nu este echivalent cu o stare conceptuală.

Dacă revenim la a doua diferență între concepte, și activități inspirate de concepte și alte proprietăți și procese, ni se prezintă o trăsătură a celor dintâi care le face în întregime rezistente la orice cercetare empirică posibilă. Un concept este, intrinsec (re)

prezențial. Conținutul lui, /lucrul/ despre care el este, îi este constitutiv. Ceva similar cu aceasta se dovedește valabil și despre celelalte trăsături. O proprietate este intrinsec atributivă, caracterul ei fiind constitutiv pentru ea. Totuși, paralela apare strict limitată, deoarece, în cazul unui concept, natura lui esențială constă în *referința la altceva*; o proprietate sau un particular. Un concept reprezintă modul în care se prezintă minții /un lucru/. Astfel, dacă ar fi posibil studiul naturii intrinseci a unui concept, acesta nu ar fi altceva decât studiul conținutului lui, adică al celui /lucru care se constituie/ drept concept. Dar întrucât un număr nelimitat de concepte pot viza același lucru – din perspective diferite – nu există o cale pe care cercetarea empirică a extensiunii conceptului s-ar putea constitui ca o cercetare științifică a conceptului însuși. Desigur, se poate considera natura conceptelor, în genere, și examina conținutul conceptelor particulare, dar acestea sunt cercetări filosofice, nu empirice.

## V

Esența celor de mai sus constă în ideea că extensionalismul pe care îl solicită atribuirea și descrierea atitudinilor psihologice se arată foarte diferit față de cel prezentat de descrierile altor fenomene non-intenționale. Elaborând – o astfel de idee pentru a face față unor posibile contraexemple, doresc să fac unele considerații asupra recentei lucrări a lui *Robert Stalnaker*, în care se încearcă elaborarea unei teorii naturaliste a intenționalității<sup>12</sup>. Autorul spune și susține că rezolvă în cadrul acestui proiect probleme similare cu obiecțiile prezentate mai sus. Comentariile mele intenționează să dea de asemenea și un răspuns (limitat) la „soluția” lui Stalnaker.

Conform punctului lui de vedere, principala problemă a intenționalității constă în modul cum se poate da o interpretare naturalistă relației dintre stările reprezentationale ale oamenilor și corespondențele (relata) acestor stări care ar fi (tot din punctul



lui de vedere) propozițiile. Stalnaker tratează această problemă în trei etape. Mai întâi își propune să arate cum pot exista relații naturale între ființele umane și entitățile abstracte non-propoziționale. În al doilea rând, să demonstreze cum pot exista relații neintenționale între persoane (și alte obiecte) și propoziții. Și, în al treilea rând, să indice cum relațiile intenționale între cei care gândesc și propoziții sunt pur și simplu materializări mai complexe ale primului tip de relații. Este posibil și necesar aici să considerăm numai partea acestui raționament, care corespunde, în principal, punctului II și III. În opoziție cu punctul de vedere conform căruia stările intenționale manifestă proprietăți care nu se pot explica naturalist. Stalnaker oferă exemple de stări non-reprezentationale care au aceleași trăsături, dar care se presupune că sunt naturale în mod neproblematic<sup>13</sup>.

Exemplele sunt: în primul rând, nevoia unui organism de a exista o anumită stare de lucruri, de exemplu, nevoia unei plante de a exista starea care presupune nitrogen în pământ. În al doilea rând, tendința unui sistem natural de a rămâne într-o stare de echilibru (dacă se află deja într-una) sau de a se reîntoarce la o astfel de stare dacă este scos din ea, de exemplu, tendința unui termostat de a menține aceeași temperatură. În al treilea rând, capacitatea unui obiect de a înmagazina informație despre mediul său, de exemplu, creșterea inelelor din trunchiul unui copac care indică o anumită vârstă. Se susține că aceste exemple includ numai stări și relații naturale, dar ele manifestă aspecte care, în mod tradițional, se presupune că aparțin doar fenomenelor intenționale și care devin, din acest motiv, inexplicabile pentru știință. Astfel, este adevărat poate, că o plantă are nevoie să obțină o anumită stare de lucruri, fără ca acest adevăr să implice și că această din urmă împrejurare va fi realizată. Lucrul de care are nevoie planta poate să nu existe, pur și simplu. De asemenea, ea poate să aibă nevoie de hrană, fără ca să existe o anumită hrană de care are nevoie. Tot așa, un sistem poate să tindă către o stare de echilibru fără să o atingă vreodată. Toate acestea vor să arate că primele două exemple crează contexte intenționale, cel puțin în măsura în care termenii relați (relata) specificați în descrițiile care îi privesc pot fi determinați și chiar ne-existenți. Al treilea tip de exemplu este oferit și cu scopul de a

prezenta un model de explicare a intenționalității diferit de acela care ar prezenta-o ca pe un exemplu de fenomen natural obișnuit, fără mistere. Proprietatea de a înmagazina informație despre mediul înconjurător, ca aceea a unui copac, sau alte modalități cunoscute, reprezintă o putere quasi-reprezentatională. Intenționalitatea, în această explicație, apare ca o varietate mai complexă de indicație naturală a unui lucru, prin altul<sup>14</sup>.

Ce să facem deci, cu aceste exemple? Toate implică ideea cauzalității și a explicației prin referință la prezența sau absența unei condiții naturale pentru ca ceva să intre într-un anumit tip de stare. Acest lucru sugerează că, probabil, non-extensionalitatea este o trăsătură comună a contextelor cauzale și deci, Stalnaker are dreptate în a susține că trăsătura atribuirii de context intențional există și în afara domeniului psihologiei. Desigur, chiar dacă am presupune că ar fi așa, numai acest lucru nu ar putea îndepărta orice piedică în elaborarea unei explicații naturaliste a minții, decât dacă ar arăta și că intensionalitatea contextelor non-psihologice ar fi neproblematică, cum presupune Stalnaker. De fapt, sunt multe de contestat în felul în care tratează el aceste probleme. Pentru început, nu este sigur că atitudinile propoziționale trebuie privite ca relații cu propozițiile. Ca explicație generală cred că o astfel de afirmație e falsă și total sortită eșecului<sup>15</sup>. În al doilea rând, pretenția lui Stalnaker că existența relațiilor între obiecte fizice și abstracte este ușor de adoptat într-o ontologie naturalistă, nu este garantată de exemplele pe care le aduce. El oferă ca exemple reprezentarea mărimilor fizice, ca înălțimea și greutatea, ca relații între purtătorii acestor proprietăți și numere. Un copac care este înalt de 3 picioare și o pisică, care cântărește 16 funzi sunt prin aceasta în legătură cu două entități abstracte: numerele 10 și respectiv 16. Adecvarea unei teorii este relativ la scopul ei explicativ. Scopul lui Stalnaker constă în a arăta cum pot sta obiectele fizice în relații reale cu entitățile abstracte fără să compromită ontologia naturalistă. Dar dacă analiza semanticii expresiilor cantitative pe care o oferă se dovedește interesantă, în această sarcină el eșuează.

„Ce se află în aceste proprietăți fizice (de tipul proprietății de a avea o anumită greutate) care face ca ele să poată fi reprezentate corect ca relații între lucrul căruia i se atribuie proprietatea

și un număr, se întreabă el. Putem înțelege astfel proprietățile – cantități fizice – deoarece aparțin unor familii de proprietăți care au o structură comună cu numerele reale. Deoarece familia de proprietăți – greutatea ale obiectelor fizice – are o astfel de structură, putem (dată fiind o unitate fixată de un obiect standard) să folosim un număr pentru a alege din familie o anumită proprietate. La aceasta cred că se reduce problema cu privire la ce înseamnă faptul că greutatea și alte cantități fizice sunt sau pot fi înțelese ca relații între obiecte fizice și numere<sup>16</sup>.

Acest citat oferă o rațiune fundamentală pentru utilizarea termenilor numerici în descrierea anumitor trăsături ale obiectelor fizice și la tratarea acestor termeni ca semnificativi din punct de vedere semantic, contribuind deci, în mod distinct, la semnificația și adevărul propozițiilor în care apar ele. Totuși, aceasta este cu totul altceva decât a arăta că astfel de termeni reprezintă relații pre-existente între obiecte și numere. Ceea ce descrie Stalnaker reprezintă o procedură prin care cantitățile pot fi caracterizate prin referința la o schemă de măsurare care împerechează sistematic proprietățile non-relaționale, fizice, cu entitățile abstracte. Ea lasă fără răspuns problema cum pot exista, dacă există, cumva, relații între înseși lucrurile fizice și cele non-fizice; deși indică felul în care predicatele – numere pot fi *aplicate* în mod corect la măsurare. Dar această explicație nu are nici o legătură cu proiectul lui Stalnaker deoarece, așa cum arată citatul, ea reprezintă o explicație în termenii puterilor *noastre* reprezentationale și a înțelegerii noastre.

Revenind la cele trei exemple, este clar că al II-lea – acela despre indicatorii naturali – se sprijină pe distincția pe care am făcut-o în secțiunea a II-a, între stările (*intenționale*) cognitive și stările *informaționale*, sau, cum le-am mai numit pe cele din urmă, *stări de sensibilitate ambientală*. În interpretarea pe care o dă exemplelor oferite (care pe lângă inelele copacului, include mercurul din termometru corelat cu temperatura ambientală, umbrele cu formele obiectelor și undele radio cu sunetele din corpul unui microfon), Stalnaker șovăie la întrebarea dacă ele implică intenționalitatea în sensul obișnuit, rezervat pentru reprezentarea mentală, dar este hotărât să susțină că relațiile cauzale de tipul celor descrise nu sunt, mai puțin reprezentationale.

Adică stările obiectelor în cauză se *reprezintă în* propoziții despre cauzele lor din mediu, totuși, concluzia nu este garantată de evidență. Relația materializată în cele câteva exemple este o relație de dependență cauzală în contexte în care starea indusă nu ar fi posibil să apară altfel. În concluzie, ea servește ca un indicator de încredere pentru altceva. Cineva care cunoaște o astfel de corelație poate să o utilizeze ca pe o sursă de informație, dar stările, ele însele, nu devin din acest motiv, reprezentative. Indiciul, informația și reprezentarea apar numai acolo unde există un subiect cognitiv angajat în interpretare sau comunicare. Pentru cine are ochi să vadă și minte să înțeleagă, iarba culcată poate "indica" pe unde a luat-o animalul în fuga lui și urma lui poate „furniza informație“ cu privire la specia respectivă. Dar fără un interpret inteligent nu există nici indiciu, nici informație, nici ceea ce presupune aceste activități, adică un conținut reprezentativ. Condiția unui fir de iarbă poate fi invocată în sprijinul unui număr infinit de ipoteze despre mediul lui, ca de exemplu că are în conținut fosfor, că este locuit de animale, că există o fabrică de ciment în apropiere, că a avut loc pășunatul săptămâna trecută, etc. Dar nu trebuie să ne batem capul să cercetăm pe care dintre stări din infinitatea lor potențială, o astfel de propoziție o reprezintă *în mod real*, deoarece nu este vorba despre niciuna dintre stările de lucru. Desigur, există o corelație între ele care poate fi exploatată cu folos de cercetători, dar aceasta este pur și simplu o relație de dependență cauzală.

Ajungem astfel la trăsătura comună a tuturor exemplelor lui Stalnaker și la sugestia că, oricare ar fi condițiile necesare pentru intenționalitate, nu putem spune că numai ea manifestă non-extensionalitate și deci, că limbajul psihologiei nu este logic independent de cel al teoriei științifice. Am făcut deja distincția între *extensionalismul pur*, care nu permite o individualizare mai fină de predicate decât cea furnizată de referința la clase de particulari, și *extensionalismul modificat*, care este sensibil la diferențe între proprietățile obiectelor. Dați-mi voie să adaug la aceasta distincția între contextele *identificării cauzale* și acelea ale *explicației cauzale*. Fără îndoială că ultimele sunt non-extensionale, și în sensul slab definibil în termenii extensionalismului pur – care se datorează nevoii de a identifica în explicație proprietățile eficace cauzale ale lucrurilor – și în sensul tare, care

este caracteristic contextelor intenționale. Adică, în contextele explicației cauzale, substituția predicatelor coexistente poate să nu păstreze valorile de adevăr nu numai când expresiile substituite nu identifică aceleași proprietăți, ci și când ele nu identifică proprietățile prin aceleași moduri de prezentare sau aceleași caracterizări conceptuale prin care au fost identificate în judecățile explicative inițiale, chiar dacă substituțiile permise sunt restrânse (ca în extensionalismul modificat) doar la predicate echivalente în mod necesar, sau la cele care semnifică aceleași proprietăți, obiective.

Și pentru a da astfel de explicații nu este suficient să se identifice corect trăsăturile relevante cauzal, deoarece există și o dimensiune euristică în acest exercițiu. Forma în care se oferă o explicație trebuie să fie inteligibilă receptorilor avuți în vedere. Conținutul ei conceptual trebuie să se încadreze (în mare măsură) în spațiul competenței cognitive a audienței. Ceea ce înseamnă că intenționalitatea este categoric prezentă în conținutul explicației cauzale. Prin contrast, în contextele identificării cauzale (care nu sunt menite a fi și explicative) nu există un astfel de element intențional. Extensionalismul pur poate fi susținut sau nu în funcție de termenii relației cauzale: dacă sunt indivizi sau proprietăți. Totuși, chiar acolo unde se poate demonstra că este inadecvat, adică acolo unde este necesar să ne referim la *trăsături* ale lucrurilor, identificarea cauzală nu implică referința de neînlăturat la modurile în care sunt prezentate gânditorilor trăsăturile obiective, la felul în care le gândesc ei (cu excepția, desigur, a situației în care cauzalitatea însăși este intențională). În consecință, judecățile cauzale de acest fel nu sunt în mod radical non-extensionale în maniera contextelor intenționale și nu reprezintă o punte între științele lumii exterioare minții și psihologie. În plus, prăpastia nu este doar logică, ci este expresia unei dezbinări ontologice în care doar de o parte se află entități intenționale ca ideile și conceptele.

Dacă ne întoarcem la primele două exemple date de Stalnaker și, care implică nevoi și tendințe non-psihologice, rămâne de arătat cauza pentru care ele nu prezintă o obiecție la teza că intenționalitatea nu este explicabilă prin naturalismul științific. Ambele tipuri de cazuri privesc relațiile cauzale între stările obiectelor și trăsăturile mediilor lor și pot crea, de aceea, o problemă pentru extensionalismul pur. Este important să vedem

de ce nu fac mai mult decât atât. Mai întâi, deci, condițiile de adevăr pentru atribuire ale acestor tipuri implică și stări de lucru intrinseci și relații între acestea. Astfel de condiții au o complexitate care nu este reprezentată în discuția despre „nevoi” și „tendințe”. Într-adevăr, foloasele aduse de economie oferă un motiv pentru a utiliza asemenea concepte teleologice în caracterizarea structurilor necesare ale relațiilor cauzale. Aici apare o întrebare în legătură cu implicațiile ontologice ale acestor atribuiri. Eu nu am o opinie formată în această problemă și consider că Stalnaker nu vrea să sugereze că ele impun recunoașterea cauzalității finale. În afară de alte neajunsuri, teoria sa nu se împacă nici cu cazurile în care obiectele nevoilor și tendințelor formulate nu există. Și el apelează tocmai la astfel de cazuri pentru a argumenta că intenționalitatea nu există *sui generis* (În orice caz, cauzele finale care nu sunt rațiuni-scop nu ar putea da naștere unui non-extensionalism tare).

Totuși, inexistența obiectului unei nevoi sau tendințe arată că, ori de câte ori este adecvat să atribuim unui sistem o astfel de stare aparent relațională, condiția de adevăr a acestei atribuiri poate fi dată în termeni care se referă doar la stări existente și relații reale. În plus, o descriere a naturii intrinseci a unei stări de altfel extrinsecă caracterizată ca o stare care are nevoie de  $p$ , sau ca fiind o consecință a lui  $c$ , sau o cauză a lui  $e$ , nu va fi nici măcar aparent relațională. Astfel, dacă este adevărat că o plantă are nevoie ca nitrogenul să fie prezent în țesuturile ei, sau că un termostat trebuie să facă în așa fel încât să se ajungă la o anumită temperatură<sup>17</sup>, atunci există descriții complete ale caracterului intrinsec al acestor stări fără referințe la cauze anterioare și la efecte posibil de realizat. Pe scurt, astfel de atribuiri teleologice ca cele prezentate de Stalnaker sunt adevărate în virtutea stărilor intrinseci și a cauzelor lor eficiente. Și se pot da descriții ale primelor care nu fac referință la alte entități. În consecință, exemplele prezentate nu au condiții de adevăr non-extensionale, de neînlăturat, exceptând probabil sensul slab care derivă din faptul că au principii mai înguste de individualizare pentru predicate decât coextensionalitatea *de facto*.

Abordând atribuirile-de-conținut psihologic, două lucruri au devenit clare. Mai întâi, că explicația adevărului lor implică

recunoașterea entităților încă mai îngust individualizate decât proprietățile naturale, așa cum au fost concepute, tradițional. Și, în al II-lea rând că, întrucât conținutul este intrinsec stărilor psihologice, nu se poate da o descriere completă a naturii lor esențiale, fără a se recurge la referința intențională, adică, referința la obiecte și proprietățile posibil non-existente. Mai puțin formal exprimat: un gând este tocmai acel ceva despre care el este și acel ceva despre care el este, poate să nu existe în realitate. Stările intenționale sunt astfel puternic, și incontestabil non-extensionale, deosebindu-se prin aceasta de obiectele cercetării științifice<sup>18</sup>.

## VI

Ca urmare a argumentelor și concluziilor expuse în cele de mai sus, apar două întrebări. Mai întâi, care ar fi relația dintre fenomenele intenționale și cele non-intenționale, sau, mai general, între minte și materie? Cea de a doua și care are legătură cu prima: care poate fi relația dintre psihologia intențională (fie ea neformală sau teoretică) și științele naturale? Argumentul meu anterior împotriva încercărilor de a explica cognitivitatea și conștiința în termeni evoluționiști a introdus ruptura între materia lipsită de spirit și conținuturile reprezentationale, într-un context temporal, pentru a demonstra imposibilitatea unei explicații naturale a apariției conștiinței de sine și a trăsăturilor mentale care o acompaniază, drept consecințe cauzale ale unor condiții anterioare, pur fizice. Legat de aceasta reiese că psihologia intenționalistă nu poate fi pur și simplu un caz special al științelor empirice de nivel inferior. Ceea ce o deosebește nu este doar faptul că ar conține concepte specializate (ca, de pildă, credință, afecțiune, înțelegere etc.) care nu apar în teoriile subsumate – așa cum poate fi adevărat și despre ecologie în raport cu biologia, biologia față de chimie și despre chimie în relație cu fizica. Mai curând, discontinuitatea radicală dintre psihologie (și alte arii de cercetare intenționalistă) și științele naturale este

marcată de non-extensionalitatea remarcabilă a primelor. Și această diferență logică, se bazează pe prăpastia ontologică dintre proprietățile fizice ale obiectelor și atributele conceptuale ale gândirii. Cu privire la problema dezvoltării istorice a intelectului pot fi puse în evidență două posibilități. *Panpsihismul* sau o formă de *mentalism emergent* care susțin: fie că trăsăturile mentale au existat dintotdeauna, dar cu timpul au devenit mai complexe, astfel că, la un anumit punct, s-a ajuns ca structurile non-mentale să manifeste proprietăți radical diferite; fie că, au fost supraadăugate, astfel că au apărut în mod misterios. Eu înclin să adopt mentalismul emergent, și, aş adăuga, nu ca o teorie, ci pur și simplu ca descriere a fenomenului care trebuie explicat. Și la acest punct înclin către ceea ce a însemnat naturalismul aristotelic așa cum a fost el prezentat de scolastica medievală.

O explicație filosofică completă a intenționalității ar indica și cum se leagă ea de constituția unor lucruri care o posedă și, de asemenea, cum este posibil că materia, anterior lipsită de intelect, a început să manifeste gânduri. Dualismul de proprietăți parcurge în ultimii ani, o perioadă de renaștere, întrucât oamenii au ajuns să recunoască inadecvarea behaviorismului și a teoriilor identității. Totuși, în același timp s-a ignorat în mare parte o problemă dificilă care ar explica cum pot fi conectate proprietăți atât de diferite categorial cu obiectul fizic căruia aceste teorii i le atribuie. Apoi, dualismul de proprietăți, în special în versiunile neo-homunculariste, preferate în mod curent, se confruntă cu dificultăți grave de ordin epistemologic – privind sorgintea conținutului cognitiv și atribuirea de trăsături mentale altor oameni. El moștenește astfel câteva dintre obiecțiunile aduse, în mod tradițional, cartezianismului clasic.

La Aquinas, și, dacă interpretez bine și la Wittgenstein, se propune o soluție mai radicală. Purtătorii stărilor psihologice poartă în același timp stări fizice. Dar ei nu sunt obiecte materiale, nu sunt, desigur, nici obiecte imateriale. Ce urmează din aceasta? Ființele umane, cum ar adăuga Aquinas, sunt animale raționale, adică, substanțe psihofizice. Această idee nu este ferită de inconveniente și are nevoie de elaborare<sup>19</sup>. Totuși, în încheiere, vreau pur și simplu s-o pun în legătură cu problema mai veche a relației dintre psihologie și științele empirice. Nu există o



posibilitate de a deriva descrierile stărilor intenționale ale persoanelor numai din descrierile obiectelor și relațiilor fizice. Această concluzie pune capăt ambiției apariției în viitor a psihologiei științifice în sensul în care a fost gândită tradițional. Ea arată și limitele explicației naturaliste – bazată pe supoziția că s-ar putea face apel doar la o ontologie materialistă. Dar acum s-ar cuveni să ne întrebăm. De ce trebuie să se facă această supoziție? Argumentele de mai sus indică falsitatea materialismului ca teorie generală a realității. Dacă ele sunt bune, atunci concluziile lor trebuie să fie eliberatoare, întrucât reasează psihologia acolo unde a plasat-o bunul simț, adică dincolo de domeniul mecanismelor fizice încurajând o concepție mai largă despre natural, care include și lucruri care nu pot fi investigate și înțelese întotdeauna drept configurații de obiecte și forțe în spațiu. Iată un drum înapoi către un punct de vedere mai vechi și mai înțelept asupra relației dintre umanism și cercetarea științifică care le vede pe ambele ca fiind preocupate de aspecte ale ordinii naturale și prin aceasta meritând în egală măsură numele de Naturalism<sup>20</sup>.

## NOTE

<sup>1</sup> Acest eseu datează din perioada unei vizite Fellowship la Centrul de Filosofie a Științei, Universitatea din Pittsburg în primăvara și vara lui 1987. Sunt recunoscător Centrului pentru invitația de a petrece timpul acolo și pentru ospitalitate. Mulțumesc de asemenea Academiei Engleze și Universității St. Andrews pentru alt sprijin financiar. O versiune a acestui eseu a fost lecturat în Stirling, Bradford, Edinburgh, Londra și Lampeter; o alta ca „an invited address” la cel de al 7-lea Simpozion de filosofie InterNordic la Reykjavik, în Mai, 1988. Sunt recunoscător și acolo gazdelor mele pentru găzduire generoasă. Textul prezent extinde și dezvoltă temele discutate în J. Haldane, „Psychoanalysis, Cognitive Psychology and Self-Consciousness”, în P. Clark and C. Wright (ed.) *Mind, Psychoanalysis and Science* (Oxford Blackwell, 1988).

<sup>2</sup> Vezi *Summa Theologiae* I-a, 1976 și *Summa Contra Gentiles*, II, 68–72.

<sup>3</sup> „Discurs asupra Metodei”, IV, în M. Haldane and G. Ross (trans), *The Philosophical Works of Descartes* (Cambridge, Cambridge University Press 1912), vol. 1, p. 101.

<sup>4</sup> Din discuțiile cu J. Miller în J. Miller, *States of Mind* (Londra BBC Publication, 1983) p. 94–95.

<sup>5</sup> Continuarea discuției asupra reflexivității și a distincției între aceasta și introspecție se află în Haldane, „Psychoanalysis, Cognitive Psychology and Self-Consciousness“, *op. cit.*, p. 129, 134.

<sup>6</sup> Astfel, nu este sigur că substituțiile predicatelor coextensive și ale expresiilor coreferențiale cu cele care apar în raza verbelor psihologice și care înregistrează modul și conținutul conștiinței (awareness) reflexive păstrează adevărul unei specificații originare de conținut, din simplul motiv că este posibil ca ele să nu capteze caracterul stării subiectului. Interferența existențială din asemenea contexte poate astfel eșua deoarece nu răspunde la conținutul experienței.

<sup>7</sup> Vezi „Intentionality and the Mental“, în H. Feigl et al. (eds.), *Minnesota Studies in the Philosophy of Science*, vol. 2 (Minneapolis: University of Minnesota Press, 1958).

<sup>8</sup> Vezi „Psychology in Physical Language“ în A. J. Ayer (ed.), *Logical Positivism* (Glencoe, Illinois: Free Press, 1959).

<sup>9</sup> Vezi „Explanation in Science and History“ în R. G. Colodny (ed.), *Frontiers of Science and Philosophy* (Pittsburg: University of Pittsburg Press, 1962; reprinted New York: University Press of America, 1983).

<sup>10</sup> În această problemă s-a scris mult și s-ar putea scrie încă. Pentru o dezbateră interesantă scrisă înainte de aflulul de publicații privitoare la ideea că gândirea implică lumea, vezi L. Chipman, „Psychological Verbs and Referential Attitudes“, în *Philosophical Quarterly* 31 (1981).

<sup>11</sup> Pentru sugestii în această problemă vezi P. Achinstein „The Identity of Properties“ în *American Philosophical Quarterly* 2 (1974); S. Shoemaker, *Identity, Cause and Mind* (Cambridge: Cambridge University Press, 1984); and R. Swinburne, *The Evolution of the Soul* (Oxford: Clarendon Press, 1986), p. 46–51. Ultimul articol prezintă o critică a teoriilor cauzale avansate în primele două studii și oferă un criteriu semantic al identității proprietății, i. e. unul intensional.

<sup>12</sup> *Inquiry* (Cambridge, Mass: MIT Press, 1984).

<sup>13</sup> *Ibid.*, cap. I, „The Problem of Intentionality“.

<sup>14</sup> Relația de indicare este definită astfel de Stalnaker: „obiectul indică P dacă și numai dacă, pentru un *a* din setul relevant al stărilor alternative ale obiectului, mai întâi obiectul se află în starea *a*, și apoi, propoziția că mediul se află în starea *f* (*a*) antrenează *p*“ (*Ibid.* p. 13). Pentru temerile utilizării acestei noțiuni într-o interpretare non-circulară a intenționalității, vezi S. Schiffer, Stalnaker's „Problem of Intentionality“, în *Pacific Philosophical Quarterly* 67 (1986) p. 92–96.

<sup>15</sup> Pentru o discuție asupra defectelor acestor puncte de vedere și posibilitatea unei alternative, vezi J. Haldane; „Brentano's Problem“, *Grazer Philosophische Studien* 34 (1989).

<sup>16</sup> „Inquiry“, *op. cit.*, p. 9.

<sup>17</sup> Formulările curioase ale acestor exemple sunt menite să se acomodeze cu sugestia lui Stalnaker că termenii acestor stări sunt propoziționali.

<sup>18</sup> Am observat înainte că răspunsurile la ideea că intensionalitatea atribuirilor de conținut semnifică rezistența intenționalității la explicația științifică poate porni de la sugestia că naturalismul este deja constrâns la non-extensionalitate, independent de orice încercare de a acomoda fenomene psihologice. Ca exemple, erau date modalitatea, judecățile despre legile naturale și cauzalitatea în genere. Cea din urmă este acum tratată în detaliu, așa cum este necesar. În ce privește primele două exemple, se cuvine un răspuns scurt. Mai întâi, formularea legilor științifice este posibil să nu necesite utilizarea modalităților *aleitice*

deoarece capacitatea legilor de a susține / propozițiile / contrafactice este explicabilă în maniera preferată de Ramsey și (David) Lewis (adică, legile pot fi privite drept consecințe logice ale acelor propoziții pe care ar trebui să le luăm ca axiome, dacă am cunoaște tot ce este de cunoscut și le-am organiza într-un sistem deductiv). Apoi, în orice caz, realismul modal nu generează intensionalitate, deoarece necesitatea este *de ce* și aceasta este modalitatea extensională. Modalitatea *de dicto*, prin contrast, reprezintă într-adevăr o problemă pentru naturalistul științific, dar numai întrucât este un caz special al intenționalității psihologicului care apare din caracterizările noastre conceptuale. Sinonimia de exemplu, este tocmai echivalența intențională sau asemănarea modului de prezentare conceptual.

<sup>19</sup> Pentru continuarea discutării acestei probleme așa cum apare ea în legătură cu natura acțiunii – de fapt problema agent-corp, vezi J. Haldane, „Folk Psychology and the Explanation of Human Behaviour”, în *Proceedings of the Aristotelian Society Suppl.* Vol. 62 (1988).

<sup>20</sup> Sunt recunoscător pentru sfaturile multor persoane și în special, ale lui Adolf Grünbaum, Nicholas Rescher și Moris Eagle care m-au încurajat să extind implicațiile argumentului împotriva psihologiei științifice. Și articolul citat anterior: „Psychoanalysis, Cognitive Psychology and Self-Consciousness” și un studiu nepublicat intitulat „Intentionality and Consciousness” (versiuni ale lor au fost citite la Scottish Philosophical Club și la Universitatea din Rice, Pittsburg și York, Toronto). S-au prezentat și au apărut variante ale acestui argument într-o formă care a putut face ca unii să creadă că eu am făcut o breșă în structura explicației ce ar putea fi umplută doar apelând la surse supranaturale – mai precis cele teologice. Răspunsul nefiresc este sugestia că naturalismul nu trebuie identificat cu știința pozitivă și dacă este dezvoltat în mod adecvat se poate explica în cadrul lui intenționalitatea și alte fenomene până în prezent problematice. Sper să apară însă clar cititorilor eseului de față că nu exclud nici cea de a doua posibilitate. Dar trebuie să fie clar, în același timp, că este o posibilitate incompatibilă cu fizicalismul. Aceasta nu înseamnă numai că explicațiile psihologice și conceptele mentale nu pot fi reduse la acelea ale fizicii, ci că ontologia corespunzătoare includ fenomenele non-fizice. În ce privește ușoara iritare cu privire la teologie, cei menționați nu greșesc presupunând că eu găsesc în existența ființelor raționale (care se aseamănă prin a avea capacitatea gândirii conceptuale, dar care diferă în perspectivele specifice ale fiecăruia) ceva care încurajează speculația teologică. Ceea ce nu face parte totuși din argumentul dezvoltat în prezentul studiu.

· Traducere Maria VORNICU și Angela BOTEZ

# EXPLICAȚIA ȘI INTERPRETAREA ACȚIUNII \*

Lars BERGSTROM

(Universitatea din Stockholm)

## 1. EXPLICAȚIILE PRIN MOTIVARE RAȚIONALĂ\*\*

De multă vreme, dar în special în ultima jumătate a secolului al 19-lea, filosofii, istoricii și oamenii de știință au discutat problema dacă explicațiile acțiunilor umane individuale (în istorie și psihologie) sunt și trebuie să fie, sub aspectele lor principale diferite de explicațiile evenimentelor fizice din științele naturale. Această discuție s-a referit uneori la problema explicației și înțelegerii, la controversa *Erklären-Verstehen*, controversă care rămâne încă foarte aprinsă; literatura din domeniu fiind vastă și în creștere. Nu se întrevăd semne de consens în viitor.

Acțiunile sunt explicate în moduri diferite. În unele cazuri explicăm o acțiune arătând unele dintre cauzele externe ale ei, ca de exemplu, atunci când spunem că cineva a comis un viol deoarece, copil fiind, a fost tratat într-un anumit fel de mama sa. În alte cazuri, explicăm *cum* s-a făcut o acțiune descriind-o în detaliu, ca de exemplu, atunci când spunem că hoțul a intrat în casă utilizând un șperaclu, pentru a deschide ușa din față. În altele, și, probabil, de cele mai multe ori, explicăm o acțiune descriind motivele pe care le-a avut agentul când a efectuat-o. Explicațiile de acest tip pot fi numite „*explicații-rațiuni*“.

Toată lumea este, în genere, de acord că explicațiile-rațiuni apar, și pot fi explicații perfecte cel puțin în viața obișnuită și în științele umaniste. Mai mult, că rațiunile la care se face (mai mult sau mai puțin explicit) referință în tipul acesta de explicații, constau

---

\* „Explanation and Interpretation of Action“ în *International Studies in the Philosophy of Science* vol. 4 nr. 1/1990.

\*\* *Reason explanation* – a fost tradus cu trei expresii care acoperă toate, după părerea noastră, sensurile sintagmei englezești: „explicații prin motivare rațională“, „explicații-rațiune“, „explicații-motiv“ (n.t.)

din credințe și dorințe (sau atitudini-pro). O explicație-rațiune tip poate suna aproximativ astfel: „Spărgătorul a intrat în casă, deoarece a dorit un loc de dormit“, sau, mai explicit: „Spărgătorul a intrat în casă, deoarece a căutat un loc de dormit și a crezut că intrând cu forța în casă ar găsi într-adevăr un loc de dormit“. Cu alte cuvinte, o „explicație – rațiune“ indică motivul pentru care s-a făcut acțiunea.

O „explicație-motiv“ poate face să crească înțelegerea de către noi a acțiunii. Același lucru este adevărat și pentru explicațiile în termenii cauzelor externe pentru explicațiile date prin descriții mai detaliate. Nici un participant la dezbateră problemei explicație-înțelegere nu ar nega (și nu ar trebui să nege) acest lucru.

Altă observație este că, ceea ce par ar fi explicațiile acțiunii în termenii cauzelor externe sunt mai degrabă explicații ale motivelor (sau ale credințelor sau dorințelor) decât ale acțiunilor. De exemplu, putem spune că spărgătorul a intrat în casă deoarece soția lui l-a încuiat afară și nu a avut unde să doarmă. Apoi, putem spune că el a spart casa deoarece cineva i-a spus că este goală. Dar, probabil, că lucrul cel mai rezonabil de spus aici este că prima dintre aceste explicații arată dorința hoțului, iar a doua credința lui și că ceea ce explică acțiunea lui, (adică motivul ei) este mai curând o combinație a acestor două explicații decât a cauzelor externe lor. De aceea „explicațiile-motiv“ par să aibă de jucat un rol central în legătură cu acțiunile.

## **2. POZITIVISM ȘI HERMENEUTICĂ**

Controversa explicație-înțelegere poate naște polemici diferite, dar, cel puțin în parte, ea se ocupă de statutul logic sau metodologic al explicațiilor-motiv. În mare, există aici două poziții în conflict. Pentru scopurile noastre ele pot fi numite poziția „pozitivistă“ și poziția „hermeneutică“. Pozitiviștii susțin că toate explicațiile – incluzând „explicațiile-motiv“ – au în comun o anumită structură logică. Această structură este descrisă de J. Stuart Mill după cum urmează: „Spunem că un fapt

individual este explicat prin arătarea cauzei lui adică, prin formularea legii sau a legilor cauzale, pentru care producerea lui este o dovadă. Și, în mod similar, spunem că o lege sau o regularitate a naturii este explicabilă, atunci când este indicată o altă lege sau legi, pentru care acea lege însăși este doar un caz și din care poate fi dedusă (1862, Cartea 3, Cap. 12, Par. 1)“.

Această teorie este deseori numită teoria explicație prin „legi de acoperire“ (*covering law*). În secolul nostru ea a fost dezvoltată și apărută în primul rând de Carl G. Hempel. În particular, deci, pozitiviștii (în sensul meu) cred că „explicațiile-motiv“ sunt cauzale și nomologice în sensul indicat de Mill și Hempel. De exemplu, Hempel spune că „explicațiile date prin rațiunile care motivează au un caracter larg dispozițional“ și că, de aceea, „ele se conformează concepției generale despre o explicație ca subsumare a *explanandumului* legilor de acoperire (legile pot avea formă strict universală sau statistică și subsumarea va fi, în consecință, deductivă sau inductiv-probabilistică) (1962, p. 18, de asemenea 1965, p. 487)“.

Potrivit poziției *hermeneutice* trebuie să facem o distincție metodologică importantă între științele naturii, pe de o parte, și științele sociale și umaniste, pe de altă parte. Din acest punct de vedere, științele naturii sunt „nomotetice“ (preocupate de legi) iar științele umaniste sunt „ideografice“ (preocupate de cazuri individuale), de aceea teoria explicației prin „legi de acoperire“ (subsumare la legi) este aplicabilă doar în științele naturii, nu și în științele umaniste. Protagonisții cunoscuți ai pozițiilor *hermeneutice* (în sensul meu) sunt Wilhelm Dilthey, William H. Dray, G. H. von Wright și Karl-Otto Apel<sup>1</sup>. Conform punctului de vedere *hermeneutic*, „explicațiile-motiv“ sunt similare cu interpretările textelor, ele explică sensul acțiunilor; ele nu sunt nici nomologice nici cauzale, și furnizează mai curând înțelegerea (*Verstehen*) decât explicația (*Erklären*).

În ultimii aproximativ 25 de ani, o mare parte din dezbaterile filosofice privind „explicația-motiv“ a fost influențată de lucrările lui Donald Davidson. Se pare că Davidson ocupă o poziție de mijloc între pozitivism și *hermeneutică*. El susține, pe de o parte, că „explicațiile-motiv“ sunt cauzale și, (deci) nomologice (1980, pp. 3, 262); iar, pe de altă parte, că există o diferență ireductibilă

între explicațiile psihologice care implică atitudinile propoziționale și explicațiile din științe ca fizica și fiziologia (1987, p. 35) Davidson crede (1980, p. 262) despre concepția lui că ar fi foarte apropiată de cea a lui Hempel. Într-o oarecare măsură această părere este adevărată deși, după cum observă un comentator (Kim, 1985, p. 383) psihologia, așa cum o înfățișează Davidson, apare mai curând „ca o cercetare hermeneutică decât ca o știință predictivă“.

Acest lucru ar putea sugera faptul că diferența dintre pozitivism și hermeneutică nu se dovedește atât de dramatică cum apare la prima vedere. În ceea ce urmează, voi încerca să arăt că este posibilă chiar o anumită reconciliere.

### 3. RAȚIUNILE ÎNȚELEGERII

Punctul de vedere pozitivist, după care „explicațiile-motiv“ sunt nomologice a fost atacat în două feluri. Un tip de obiecție este acela că acțiunile umane *nu sunt guvernate de legi*, cel puțin nu de legi exacte. Voi reveni la această idee în secțiunea următoare. O altă obiecție este că legile și relațiile cauzale sunt pur și simplu *irrelevante* în contextul explicațiilor – prin motivare rațională, adică, din punct de vedere al științelor umaniste (incluzând științele sociale hermeneutice). Voi încerca acum să arăt că această a doua obiecție nu este o obiecție care se susține.

Obiecția pe care o am în minte postulează faptul că științele naturale și științele umaniste au interese principale de cunoaștere“ diferite. Se spune că științele naturii vizează explicația cauzală, predicția, controlul etc., în timp ce științele umaniste au ca scop interpretarea, comunicarea și înțelegerea intersubiectivă. S-ar părea că Dray are în vedere tocmai această diferență atunci când scrie: „Principală mea obiecție la acceptarea doctrinei legilor de acoperire în istorie nu este dificultatea de a opera cu ele fie în forma total deductivă fie în cea parțială, ci mai curând faptul că ea fixează un fel de *barieră conceptuală* pentru o istoriografie orientată umanist (1963, p. 133)“.

Presupusa diferență se poate ilustra cel mai bine printr-un exemplu: Să considerăm uciderea lui Cezar de către Brutus. Este posibil să ne intereseze ceea ce s-a întâmplat în mintea lui Brutus chiar înainte de a-l înjunghia pe Cezar. Care erau credințele și dorințele lui? Apoi este posibil să vrem să știm ce credea Brutus însuși că ar fi motivul lui pentru a-l ucide pe Cezar (în măsura în care el avea o idee clară despre acesta). Poate că am dori să cunoaștem și dacă acțiunea lui Brutus era rațională, date fiind credințele și dorințele lui la acea dată. Dacă am avea un răspuns la toate aceste întrebări am ști foarte mult despre Brutus. Într-un sens, am putea afla și care era „motivul lui”. Deocamdată nu știm însă dacă acela a fost motivul adevărat care l-a făcut pe Brutus să acționeze așa cum a acționat. Nu știm dacă Brutus a făcut ce a făcut *din cauza* respectivă. Deci, nu avem încă o explicație a acțiunii lui Brutus.

Să presupunem că Brutus dorea reinstaurarea unei guvernări republicane la Roma și că el credea că moartea lui Cezar ar duce la împlinirea dorinței. Să presupunem că aceasta era și ceea ce gândea Brutus însuși despre motivul său de a-l ucide pe Cezar. Totuși nu urmează de aici că Brutus a făcut ce a făcut deoarece avea această dorință și această credință. Nu putem presupune că Brutus era un expert infailibil în ceea ce privește propriile lui motive. (Majoritatea dintre noi nu suntem). Poate motivul lui real era dorința lui de a obține mai multă putere și credința lui era că uciderea lui Cezar era un mijloc pentru a atinge acest scop. Sau poate pur și simplu îl ura pe Cezar și dorea să-l vadă mort.

Dacă nu mă înșel, susținătorii tradiției hermeneutice ar fi de părere că: până acum deținem înțelegere, dar nici o explicație. Dar ei ar merge un pas mai departe, și ar pretinde că din punct de vedere hermeneutic numai de înțelegere (*Verstehen*) avem nevoie. Explicația (*Eklären*) se află dincolo de scopul nostru. Explicația are de a face cu cauzalitate și legi, dar acest lucru pur și simplu nu interesează din punct de vedere hermeneutic.

Poziția pare stranie. Desigur, este un truism faptul că dacă nu te interesează explicațiile, nu te interesează motivele. Totuși, dacă te interesează atât de mult ce s-a întâmplat în cazul lui Brutus în momentul crimei, este ciudat să fii indiferent la cunoașterea motivului pentru care a comis el crima (motiv care se deosebește de propriul său răspuns la întrebarea asta, dacă a avut unul).



În plus, se pare că *trebuie* să te intereseze și explicația, dacă e interesează înțelegerea (*Verstehen*), interpretarea. Deoarece e pare că nu ai nici o bază pentru a atribui credințe și dorințe unei persoane, decât dacă aceste credințe și dorințe explică niște fapte sau o dovadă pe care o ai. În cazul lui Brutus, aceste date pot fi frazele pe care se spune că le-a spus sau scrisorile pe care le-a scris lui Cezar etc. Vrei să găsești o interpretare care explică toate datele de care dispui cât se poate de bine, în consecință, înțelegerea apare drept rezultat al procesului de interpretare care nu este altceva decât obișnuita (*inference to the best explanation*) „ineferență la cea mai bună explicație“. Deci înțelegerea presupune explicația. Contrar la ceea ce se pretinde deseori, *Verstehen* presupune și implică *Erklären*. Pe de altă parte, și *Erklären* pare să presupună și să implice *Verstehen*. Deoarece datele care sprijină o interpretare nu pot fi deduse, formal, din ea. Mai curând, relația se prezintă astfel: informațiile sunt „lucrul la care te-ai aștepta“ sau „ceea ce ar fi firesc“, dacă interpretarea ar fi corectă. În cazul lui Brutus, am putea raționa astfel: dacă el ar fi dorit să reinstaureze republica, frazele sale ar trebui să conțină o judecată în acest sens; dacă da, și nu există probe contrarii, (probabil) că el asta a dorit. Dar pentru a ști la ce să ne așteptăm, dat fiind că o anumită interpretare este corectă, trebuie să utilizăm un fel de empatie sau *Verstehen* trebuie să înțelegem cum s-ar manifesta o credință dată sau o atitudine în diverse condiții. De exemplu ar trebui să ne imaginăm că suntem în situația lui Brutus și să ne imaginăm ce i-am fi spus lui Casius și altor oameni și ce altceva am mai fi făcut.

Sușținătorii lui *Verstehen* au dreptate deci, când zic că *Verstehen* are un loc nu numai în contextul descoperirii (așa cum recunosc Hempel și alți pozitivisti), ci și în contextul justificării<sup>2</sup>.

În unele situații, credințele și dorințele se pot manifesta în stări și evenimente observabile care nu sunt acțiuni (de exemplu într-o îmbujorare sau într-un tremur). Dar mult mai mult se manifestă în acțiuni. În general, deci, înțelegem pe oameni în măsura în care putem da explicații rezultate prin motivarea rațională a acțiunilor lor. În plus, dat fiind că simptomele sunt legate cauzal cu lucrul pentru care sunt simptome, se pare că aceste explicații-rațiuni trebuie să fie cauzale și deci, nomologice. Este ceea ce suțin pozitivisti<sup>2</sup>.

#### 4. ACȚIUNI GUVERNATE DE LEGI

Totuși, așa cum am spus, mai există o obiecție plauzibilă la punctul de vedere pozitivist că explicațiile prin motivare rațională sunt nomologice, și anume că acțiunile umane nu sunt guvernate de legi. Această obiecție are câteva versiuni.

(1) O versiune pornește de la supoziția că noi suntem liberi și ca atare am fi putut acționa altfel, deci acțiunile noastre nu sunt guvernate de legi naturale. Nu vreau să mă refer aici la vechea problemă dacă libertatea este compatibilă cu determinismul. Totuși, se poate susține că „explicațiile-motiv“ sunt pur și simplu imposibile dacă acțiunile nu sunt guvernate de legi. Dacă dorințele și credințele lui Brutus nu sunt legate logic de acțiunea sa atunci ele nu ne furnizează nici un temei pentru a ne aștepta ca acțiunea să se producă, ele nu explică acțiunea. Problema de ce a apărut acțiunea rămâne deschisă.

Desigur, argumentația de mai sus demonstrează că pozitivistii ar avea dreptate în problema explicațiilor-motiv. Se pare însă că acestea nu explică acțiunile în realitatea lor. Anti-pozitivistii pretind că „explicația-motiv“ nu explică o acțiune în sensul de a arăta că acțiunea era de așteptat, ci doar descrie starea mentală a agentului la momentul acțiunii. Rămâne neclar cum o astfel de descriere, dacă nu explică acțiunea, poate să sporească înțelegerea ei de către noi. Ea poate ajuta la înțelegerea agentului, dar nu a acțiunii.

(2) O versiune diferită a obiecției este că „omul nu are o natură generală“ în sensul că nu există legi universale valabile care să guverneze toate acțiunile lui. Legile sociale au caracter istoric, ceea ce înseamnă că ele sunt valide doar pentru societăți particulare în perioade de timp specifice<sup>3</sup>. Nu este însă suficient să arăți că teoria legii de acoperire este inaplicabilă în cazul acțiunilor umane pentru a nega existența oricăror legi. Aceasta arată cel mult că legile implicate în explicațiile-motiv ar trebui să fie „istorice“.

(3) Altă versiune spune că acțiunile umane sunt unice și deci nu pot fi privite ca exemple de structuri regulate de comportament care ar putea fi descrise prin legi. Împotriva acestei poziții se subliniază deseori că toate fenomenele individuale, inclusiv

venimentele fizice, sunt unice. Nu pot exista două fenomene care să fie absolut la fel din toate punctele de vedere. Dimpotrivă, două evenimente sunt similare în unele privințe și diferite în altele. Legile nu cer similaritate deplină, mai curând ele spun sau implică că evenimentele similare într-un anumit mod sunt similare și în alte moduri. Acest lucru s-ar putea să fie adevărat și în cazul legilor privitoare la acțiuni.

(4) Versiuni mai sofisticate ale obiecției pretind că dacă există anumite „legi” care leagă motivele și acțiunile, acestea nu sunt legi clasice riguroase și exacte.

Problema poate fi ilustrată cu ajutorul unui exemplu. Să presupune că acțiunea lui Brutus de înjunghiere a lui Cezar este explicată prin referința la dorința lui de a-l ucide pe Cezar și prin credința lui că înjunghierea lui Cezar este un mod de a-l ucide. Dacă un teoretician al legii de acoperire ar detalia această explicație el ar ajunge la un raționament asemănător cu cel de mai jos:

1. Brutus a dorit să-l omoare pe Cezar.
2. Brutus a crezut că înjunghiindu-l îl va ucide.
3. Brutus a fost în împrejurarea de tip C.
4. Oricine este în împrejurarea de tip C și care vrea să realizeze ceva X și care crede că făcând Y e un mod de a realiza X, va face Y
5. Deci, Brutus l-a înjunghiat pe Cezar.

Ceea ce avem aici este o inferență în care concluzia urmează din premise și unde niște premise (1, 2 și 3) formulează „condițiile inițiale” și o premiză (4) are aparența unei legi generale. Așa ar putea arăta o explicație conform modelului lege de acoperire. Mai mult, pentru ca explicația să fie corectă, premisele trebuie să fie adevărate. De aceea avem nevoie de o referință la împrejurări specifice C, deoarece, în mod clar, nu este adevărat că oricine ar fi cel care dorește X și crede că a face Y este o cale de a realiza X va face Y. De exemplu, Y poate fi prea costisitor. Totuși, până acum C nu a fost specificat. Pentru scopurile modelului legii de acoperire e nevoie să specificăm tocmai o condiție C care este astfel încât:

- a) face adevărate premisele (3) și (4),
- b) face ca premiza (4) să fie asemănătoare unei legi,
- c) dă conținut empiric premisei.

Acum problema este dacă poate să *nu existe* nici o condiție C care să satisfacă simultan toate aceste cerințe. (În orice caz, este foarte sigur că nu putem ști niciodată cum ar arăta în detaliu o astfel de condiție C). Probabil că C, ar trebui să includă ceva de felul următor: agentul crede că poate face Y și, nu există o alternativă la Y despre care el să creadă că ar fi la fel de eficientă sau mai eficientă în satisfacerea dorințelor lui. Dar doar atât nu este destul. Premisa (4) poate fi falsă deoarece în unele cazuri agentul pur și simplu leșină sau moare înainte de a avea timp să producă acțiunea. Și dacă astfel de cazuri nu se întâmplă și (4) devine de fapt adevărat, aceasta pare să fie numai din întâmplare. Deci (4) nu poate fi asemănătoare legii.

Apoi, dacă C este înlocuit de condiția *ceteris paribus* „alte lucruri fiind neschimbate“ atunci (4) nu ar avea un conținut empiric clar. În cazul în care credința și dorința sunt prezente dar acțiunea nu apare, am putea spune că lucrurile nu au stat la fel. La fel s-ar petrece lucrurile dacă am reuși să detaliem o condiție foarte complicată care înregistrează negația a orice s-ar putea concepe că merge rău și care astfel îl face pe (4) să înlăture orice contraexemplu posibil. În ambele cazuri noi l-am fi făcut pe (4) imun la revizie și astfel l-am fi lipsit de conținut empiric și de valoarea explicativă.

Există cel puțin două modalități în care am putea încerca să tratăm această dificultate. Mai întâi, am putea încerca să înlocuim premiza (4) printr-o lege *statistică*. În acest caz, alegerea lui C poate să nu pară a fi o problemă așa de mare. Putem pur și simplu să alegem pe C astfel încât să existe o probabilitate mare să fie adevărat consecventul lui 4 când antecedentul este adevărat. Totuși, dacă noua noastră premiză statistică specifică o probabilitate numerică avem o problemă asemănătoare cu cea dinainte: este greu să ne dăm seama cum putem obține și asemănarea cu legea și adevărul. Dacă (4) nu dă o probabilitate precisă, dar conține un calificativ ca „în majoritatea cazurilor“ atunci nu are un conținut clar și poate nici să nu semene cu o lege.

În al doilea rând, se poate pretinde că premisele 1 și 2 sunt suficiente să explice 5, dată fiind o specificare plauzibilă a lui 3, sau dată fiind existența unei condiții C care să satisfacă 3 și 4, deoarece, în afară de a fi adevărate și empirice, ele sunt asemănătoare legii. În acest caz, nu contează dacă premiza (4) este o lege empirică, deoarece conținutul empiric al explicației, ca și caracterul ei nomologic poate fi conținut în 1 și 2 mai curând decât în (4). Punctul principal apare deoarece atribuirea de credințe și dorințe implică dispoziții (sau tendințe sau propensități) care se vor manifesta în, sau vor cauza, anumite tipuri de acțiuni în anumite condiții. Acesta nu este tocmai felul în care Hempel a dorit să aplice modelul lege de acoperire la explicațiile prin motivare rațională dar este ceea ce sugerează de exemplu Davidson, care pretinde că punctul lui de vedere este foarte apropiat de cel al lui Hempel. Davidson scrie: „Hempel a început prin a arăta că explicațiile-rațiuni nu diferă prin caracterul lor logic general, de explicațiile din fizică sau din alte domenii. Meditațiile mele întăresc acest punct de vedere. Totuși într-o problemă nu sunt convins: legile implicate în explicațiile-rațiuni mi se pare că privesc doar individuali, ele sunt generalizările prezente în atribuiri de atitudini, credințe și trăsături (1980, p. 274)“.

Am sugerat că Davidson ocupă o poziție de mijloc între pozitivism și hermeneutică. Citatul de mai sus ne sprijină părerea, deoarece el arată că Davidson privește explicațiile-rațiuni ca fiind, și nomotetice și ideografice; ele implică legi, dar se ocupă de cazuri individuale. Totuși, Davidson, mai susține și că legile implicate în explicațiile-rațiuni nu sunt cele mai potrivite, ele nu sunt destul de „grave“ sau „stricte“ și în consecință, sunt insuficiente pentru explicație. Voi cerceta acum această idee.

## 5. LEGI ȘI CAUZALITATE

Potrivit lui Davidson, explicațiile-rațiuni sunt cauzale. Motivele explică acțiunile doar dacă le cauzează. Pe de altă parte, Davidson afirmă însă că relațiile cauzale presupun legi *stricte* și

că legile psihologice și legile psihofizice nu pot fi stricte. (Legile psihologice nu pot fi stricte deoarece mentalul nu este un sistem închis, legile psihofizice nu pot fi stricte deoarece există aici un element normativ ireductibil în atribuirea atitudinilor care nu joacă nici un rol în fizică). Legile care sunt implicite în explicațiile-motiv sunt fie psihologice fie psihofizice. Ele sunt generalizări aproximative care nu pot fi modelate în legi stricte. Urmează că ele prin ele însele, nu reușesc să susțină relațiile cauzale. De aceea, dacă motivele cauzează acțiuni – și, în consecință, dacă motivele explică acțiuni – aceasta poate fi numai în virtutea altor legi, adică a legilor fizice<sup>4</sup>.

Aici am două întrebări. Mai întâi, cum știm că relațiile cauzale între motive și acțiuni cer legi stricte mai curând, decât legi non-stricte? De ce nu sunt suficiente legile psihologice aproximative care, potrivit lui Davidson, sunt implicite în explicațiile-motiv? A doua, cum știm că există, într-adevăr, legi stricte care leagă motivele cu acțiunile, în cazurile în care credem că avem o explicație-motiv corectă?

Desigur că există și alte posibilități. Una ar fi să admitem că deși cere legi, cauzalitatea, nu cere neapărat legi stricte. Alta este că ar exista două feluri de cauzalitate: strictă și non-strictă. (corespunzând tipului de lege implicat) și că motivele cauzează acțiunile doar în sens non-strict.

Dagfin Follesdal observă că Davidson nu ne-a dat nici un argument pentru punctul de vedere că relațiile cauzale ar cere legi stricte (1985, p. 318). Și este greu să vedem cum ar arăta un astfel de argument. Cineva ar putea pretinde, urmându-l pe Hume, că toate cazurile în care credem că un eveniment îl cauzează pe altul reprezintă o lege strictă care acoperă cazul dat. Apare clar că lucrurile nu stau așa când credem că un motiv cauzează o acțiune. Deoarece în astfel de cazuri noi nu am găsit o lege strictă care să subsumeze cazul. Noi nu știm și nici măcar nu putem ghici care ar fi legea relevantă, dacă există vreuna. De aceea, nici nu putem crede că o anumită pereche de evenimente reprezintă o materializare a legii. Davidson ar putea spune că pentru început am găsit doar generalizări aproximative, dar că în unele domenii, în special în cadrul științelor naturii, am fost în stare să le transformăm în legi stricte. Probabil este adevărat. Dar

chiar așa fiind nu putem conchide că oriunde există cauzalitate (sau oriunde e o generalizare aproximativă este și aproximativ adevărată) există la îndemână o lege strictă care subsumează cazul. Ca exemplu de inducție afirmația se dovedește prea riscantă.

Davidson susține că s-ar putea stabili „standarde foarte ridicate” pentru ceea ce el numește „legi stricte” și anume să nu conțină propoziții „Ceteris paribus” și „să ajungă atât de exacte, încât să permită predicția necondiționată a evenimentului în cauză, tot așa cum permite caracterul probabilistic ireductibil al fizicii (1987, p. 44). Perfect, Davidson e liber să folosească expresia „lege strictă” cum vrea. Dar el mai spune și că generalizările aproximative furnizează probe pentru existența legilor stricte (1980, p. 16). Aceasta apare însă deja cam îndoielnic. Pentru a explica generalizările nu trebuie să presupunem nici o lege strictă. Este absolut suficient să presupun că există anumite puteri cauzale, sau relații de tipul legilor între universalii (adică legi non-stricte) care se manifestă în moduri care pot fi descrise prin generalizările respective.

Totuși, dacă nu există legi stricte care să conecteze motivele și acțiunile, nu înseamnă că motivele nu cauzează acțiunile. Nu înseamnă nici că modelul-lege de acoperire este nepotrivit în cazul explicațiilor-rațiuni. De aceea nu este evident că modelul-lege de acoperire și cauzalitatea impun legi care apar stricte în sensul lui Davidson.

O ilustrare simplă poate fi folositoare aici. Să considerăm un calculator electronic: numerele expuse sunt cauzal determinate de butoanele pe care se apasă, deci calculatorul nu este un sistem închis deoarece factori din afara lui determină care butoane să fie apăsată și, pe lângă aceasta, care ar putea face să se închidă curentul sau să se dezintegreze calculatorul întrerupându-se legătura între butoanele pe care se apasă și numerele expuse<sup>5</sup>. Astfel, nu mai apare ca lege strictă, (în sensul lui Davidson), ca ori de câte ori se apasă pe butoanele „On”, „2”, „X”, „3”, și „=” în această ordine, să apară numărul „6”. Desigur, există aici o generalizare aproximativă și pare foarte clar că tocmai presiunea pe butoane este aceea care cauzează apariția numărului „6”. Calculatorul fiind construit în așa fel încât la diferite butoane sunt atașate anumite tendințe, dispoziții sau puteri cauzale.

Similar, aş spune, că un agent uman: el sau ea, dacă are anumite credinţe şi dorinţe, va acţiona în anumite feluri, în anumite condiţii, restul rămânând neschimbat. Cu cuvintele lui Davidson: „faptul că pot cauza anumite acţiuni în condiţii adecvate reprezintă un aspect fundamental al unei credinţe sau dorinţe“ (1985, p. 246). Hempel are un punct de vedere similar. El susţine că „a atribui cuiva o anumită credinţă sau scop avut în vedere, înseamnă a implica faptul că în anumite condiţii va tinde să se comporte în anumite feluri“ (1965, p. 473).

Astfel, a avea o credinţă sau dorinţă, este analog, în cazul calculatorului, cu „a avea anumite butoane pe care se apasă“. Ceea ce se întâmplă în ambele cazuri este guvernat de legi non-stricte. Agenţii umani ca şi calculatoarele electronice sunt sisteme deschise tuturor tipurilor de perturbări venite din mediile lor. Fără îndoială, ceea ce fac ele este cauzat şi poate fi explicat prin referinţă la stările lor interne (care pot fi la rândul lor cauzate parţial de factori externi).

## 6. MOTIVUL ÎN CALITATE DE CAUZĂ

Totuşi, dacă aşa se prezintă relaţia între motive şi acţiuni, se pare că poziţia hermeneutică ar putea să fie corectă. De exemplu, dacă am dat o interpretare suficient de detaliată dorinţelor şi credinţelor lui Brutus când l-a omorât pe Cezar, incluzând, o descriere a puterilor lor relative, atunci nu trebuie adăugat nimic pentru a explica această acţiune. În particular, nu mai trebuie să adăugăm că un motiv a cauzat acţiunea, deoarece acest lucru se găseşte deja implicat în descrierea dorinţelor şi credinţelor, ce explică faptul că a apărut acţiunea.

Rămâne totuşi adevărat că este posibil ca cineva să aibă un motiv pentru a acţiona într-un anumit mod, şi nu acel motiv să cauzeze acţiunea respectivă. La o primă aproximaţie, pot exista trei feluri diferite de explicaţie a acestui lucru. Mai întâi, agentul poate avea un alt motiv, *mai puternic* (sau la fel de puternic)



pentru a performa acțiunea. Acest *alt* motiv este atunci cauza (sau principala cauză) pentru performarea acțiunii. În al doilea rând, agentul poate avea un motiv mai puternic (sau la fel de puternic) pentru a *nu* performa acțiunea. În acest caz, el poate să nu performeze. În al treilea rând, ar putea exista o barieră *externă* care se dovedește suficient de puternică pentru a-l împiedica să performeze acțiunea. Spunând că un motiv dat a cauzat acțiunea afirmăm, implicit, că nici unul dintre acești trei factori nu a fost prezent. Faptul poate fi formulat și explicit. Susținerea că nu existau alte motive mai puternice poate face parte dintr-o descriere sau interpretare hermeneutică. În cazul în care se impune o explicație a absenței barierelor externe care ar fi putut împiedica acțiunea, ea poate fi găsită în studierea contextului.

Dacă astfel de afirmații sunt corecte, ar fi posibil să acceptăm și pretenția lui Dray că o explicație-motiv (adică, în terminologia lui „o explicație rațională”) prezintă acțiunea în cauză ca fiind, într-un sens, adecvată sau justificată în împrejurările date, sau ca „un lucru de făcut pentru rațiuni date” (1957, p. 124) De exemplu, să considerăm explicația că Brutus l-a omorât pe Cezar deoarece dorea să reinstaureze republica în Roma. Explicația indică un motiv. Dacă el a fost într-adevăr cauza acțiunii lui Brutus, înseamnă că Brutus nu ar fi putut avea un motiv mai puternic nici pentru a-l omorî nici pentru a nu-l omorî pe Cezar. De aici, explicația sugerează că acțiunea lui Brutus a fost rațională sau justificată din punctul de vedere al credințelor și dorințelor lui Brutus la acea dată. (Dar, desigur nu este nevoie ca aceste credințe și dorințe să fie, la rândul lor, raționale în sine. Brutus însuși poate să fi fost irațional din anumite puncte de vedere. De exemplu, o dorință puternică a lui s-ar fi putut baza pe foarte puțină evidență). Așa că, se pare că explicația noastră cauzală bazată pe motivarea rațională a acțiunii lui Brutus poate fi în același timp o explicație rațională în sensul lui Dray.

Hempel a argumentat că o explicație rațională în sensul lui Dray nu poate arăta de ce un agent a acționat într-un anumit fel, ea nu poate decât să ofere teren pentru a crede că ar fi fost rațional ca agentul să acționeze în acest fel (1965, pp. 470–471). Aici mi se pare că Hempel greșește. Am văzut deja că o explicație cauzală bazată pe motive este în același timp rațională. În schimb,

se pare că o explicație rațională ar trebui să fie cauzală. O explicație rațională trebuie să ofere motivele agentului de a face ceea ce a făcut. Desigur, acestea trebuie să includă motivele lui cele mai puternice pentru a face ceea ce a făcut. Deoarece el a făcut ce a făcut, motivele lui cele mai puternice trebuie să fi fost eficiente cauzal. De aceea explicația este cauzală. De fapt, mi se pare că accentul lui Dray pe raționalitate nu diferă mult de punctul de vedere al lui Davidson după care atunci când un motiv cauzează o acțiune, motivul reprezintă o „cauză rațională“, ceea ce înseamnă „desemnarea cauzei“ drept o credință și o dorință în lumina căreia acțiunea apare ca rezonabilă (1980, p. 233).

## 7. GENERALITĂȚI

Să ne întoarcem, pentru moment, la diferența dintre punctele de vedere ale lui Davidson și Hempel. Davidson spune că diferența dacă există vreuna, privește tipul de legi prezente implicit în explicațiile-motiv. Potrivit propriei sale concepții, legile sunt *personale*, în sensul că ele privesc numai persoana a cărei acțiune o explicăm, în timp ce pentru Hempel, ele sunt *generale*, și privesc pe toți agenții sau pe toți agenții raționali (Davidson 1980, pp. 273–74). În termenii exemplului nostru din paragraful (4) de mai sus, am putea spune că pentru Hempel este esențial ceva de genul punctului (4) iar pentru Davidson, premiza (4) ar putea fi foarte bine îndepărtată, deoarece (1) și (2) sunt nomologice. Unii ar putea afirma că este esențial pentru modelul lege-de-acoperire, ca legile implicate să fie generale. Urmează ca acest model să fie abandonat sau măcar mult modificat de către cei care cred că ceea ce fac legile nomologice, poate fi făcut și de legile personale.

Totuși, în exemplul nostru, dacă premiza (4) poate să cadă, acest lucru se poate întâmpla numai întrucât ceva asemănător ei este inclus deja în celelalte premise. În orice caz, ceva de tipul lui (4) trebuie să fie adevărat, altfel explicația este dovedită greșită. În particular, dacă Brutus nu satisface un set de condiții, astfel încât, *oricine* ar fi acela care ar satisface acele condiții, ar acționa ca Brutus, atunci, desigur, rămâne un mister de ce Brutus a

acționat cum a acționat, ne mai fiind posibilă nici o explicație. În consecință, toate explicațiile-rațiuni par să presupună adevărul unei generalizări ca (4). (În ce măsură o astfel de generalizare trebuie să aibă forma legii și să fie empirică, rămâne un lucru controversabil).

În practică, se pare că cerința de generalitate apare în ideea, de obicei atribuită tradiției hermeneutice, că înțeleg acțiunea lui Brutus – *Verstehen* – când recunosc că aș fi făcut același lucru dacă aș fi fost în locul lui. Deoarece ceea ce recunosc nu reprezintă doar ceva comun mie și lui Brutus, rezultă că ceea ce aș fi făcut eu dacă aș fi fost în locul lui Brutus, ar fi făcut *oricine* (sau orice ființă rațională) dacă ar fi fost în locul lui. Nu este vorba despre ceva special în legătură cu mine, ci de ceva ce trebuie să fie adevărat pentru mine dacă este adevărat pentru oricine. Așa că într-o astfel de situație diferența dintre pozitivism și hermeneutică apare mult mai puțin substanțială decât se consideră de obicei.

Mai mult, nu sunt așa de sigur nici că există într-adevăr, o diferență între Davidson și Hempel. Cred că Davidson ar fi de acord că trebuie să fie (aproximativ) adevărată o generalizare ca (4), dacă motivul lui Brutus furnizează o explicație a acțiunii lui Brutus. Hempel ar înlocui pe „oricine” din (4) (sau acolo unde este specificat) cu „orice agent rațional” și ar include în C condiția ca Brutus să fie un agent rațional (1965, p. 471). Dar raționalitatea implicată se dovedește aceeași ca și raționalitatea din modelul lui Dray, fiind compatibilă și cu o cantitate considerabilă de iraționalitate. Davidson a accentuat deseori faptul că, pentru a interpreta alte persoane și a le explica acțiunile, trebuie să presupunem că sunt, în mare măsură, raționale<sup>6</sup>. Hempel are un argument similar. El mai susține și că „raționalitatea acțiunii umane nu poate fi garantată de convenții implicite în criteriile care guvernează atribuirea de scopuri și credințe agenților umani. (1965, p. 477). Sunt sigur că și Davidson și Dray ar accepta acest lucru.

Este adevărat că Davidson îi obiectează lui Hempel faptul că explicației în termenii raționalității așa cum o susține el, „i-ar lipsi forța explicativă” (1987, p. 43). Cred că el înțelege prin aceasta tocmai faptul că o premiză ca (4) din exemplul nostru (dar restrânsă la agenți raționali) nu reprezintă o generalizare empirică, ci mai curând o definiție a „raționalității”<sup>7</sup>. Aceasta ar fi o mutare posibilă pentru cineva care dorește să tragă o linie clară de demar-

cație între generalizările empirice și definiții (sau între propoziții „sintetice” și „analitice”). Totuși, nici Davidson nici Hempel nu par doritori să accepte o astfel de distincție<sup>8</sup>, de aceea și eu cred că diferența dintre ei este măcar din punctul respectiv de vedere neglijabilă.

## 8. LEGI PSIHOFIZICE

Totuși, există un alt mod în care Hempel și Davidson sunt în dezacord. După Davidson, „explicațiile – rațiune” implică legi psihofizice, legi, în mod special, problematice. Nu există nici o teză similară cu aceasta în lucrările lui Hempel așa că, probabil, aici se află punctul în care cei doi sunt într-adevăr, în dezacord. Oricum, se dovedește indicat să cercetăm mai în detaliu ideea lui Davidson despre legile psihofizice.

După cum am văzut, Davidson susține că există o „diferență ireconciliabilă” între explicațiile prin motivare rațională și explicațiile fizice. În același timp, el crede ca și Hempel, că cele două tipuri de explicație sunt similare din multe puncte de vedere fiind și cauzale și nomologice. Diferența apare în vocabular. După Davidson conceptele mentale nu pot fi reduse sau legate prin legi stricte de concepte fizice<sup>9</sup>. El scrie: „Diferența esențială care cred că există între explicațiile rațiunii și explicațiile în ultimă instanță fizice poate fi expusă astfel: legile care leagă mentalul și fizicul nu sunt la fel cu legile fizicii, și nu pot fi reduse la ele” (1987, p. 45).

Se pare că problema importantă pentru Davidson nu constă doar în faptul că legile psihofizice ar fi *non-stricte*, nici că ele ar fi *personale* (non-generale). Mai curând, se pare că deranjează faptul că ele sunt legi *mixte* în sensul că au de a face cu două vocabulare distincte, care nu se potrivesc și nu se pot potrivi unul cu altul. Iată o idee pe care nu am luat-o încă în considerație.

Vocabularul psihologic sau mentalist, relevant aici, reprezintă vocabularul pe care îl folosim pentru a atribui oamenilor acțiuni și atitudini propoziționale, precum și credințe și dorințe.

Davidson susține că acest vocabular are un caracter „*normativ*” (1987, p. 46). Înseamnă că trebuie să folosim propriile norme sau standarde privind consistența, corectitudinea și raționalitatea pentru a identifica rațiunile și acțiunile altor oameni. Pe de altă parte, vocabularul nostru fizical nu are un caracter normativ. Deci, cele două vocabulare nu se potrivesc”<sup>10</sup>.

Nu pare așa ușor să înțelegi acest lucru. S-ar putea presupune că raționamentul sună, în linii mari, astfel: (1) normele sunt necesare pentru a identifica fenomenele mentale; (2) dar dacă ar exista legi psihofizice stricte, acestea ar putea fi folosite pentru a identifica stările mentale pe bază de dovezi pur fizice, fără norme; (3) deci, nu pot exista legi strict psihofizice. Jaegwon Kim este cel care interpretează raționamentul lui Davidson în felul de mai sus (1985, pp. 378–381).

Totuși, în acest construct raționamentul nu este convingător deoarece (2) nu este adevărat. Legile psihofizice pot exista chiar dacă nu vom ști niciodată care sunt ele, dar dacă e așa, noi nu putem să le folosim pentru a identifica stările mentale. Deci, existența unor legi stricte psihofizice poate fi compatibilă cu caracterul normativ al conceptelor mentale.

Cred că avem posibilitatea să găsim o interpretare mai bună pentru raționamentul lui Davidson luând în considerare că el susține că aceasta este în legătură cu doctrina indeterminării traducerii a lui Quine (1980, p. 222, 257, 1985, pp. 245). Potrivit acesteia în cadrul unei anumite arii de posibilități diferite nu există o esență reală pe care o semnifică propoziția. Mai mult, același lucru este adevărat și despre ceea ce crede sau dorește o persoană, întrucât credințele și dorințele sunt identificate prin intermediul propozițiilor. Deci argumentul lui Davidson poate fi rezumat astfel: (1) Este o realitate că o stare fizică anume se obține într-o situație dată (2) Astfel, dacă ar exista legi psihofizice stricte, ar exista și o realitate anume de la care stările mentale să derive; (3) Dar nu există o realitate definită a stărilor mentale; (4) În consecință nu pot exista legi psihofizice. Pentru ca acest raționament să fie valid, legile menționate în (2) trebuie să fie de un anumit tip. Să presupunem, de exemplu, că există legi psihofizice, dar că toate au forma:

$$\forall_x \left[ \Psi_1 x \rightarrow (\phi x \rightarrow \Psi_2 x) \right]$$

unde  $\emptyset$  și  $\psi$  sunt predicate psihofizice și  $x$  este un predicat fizic. Lucrul pare posibil dacă mentalul este un sistem deschis afectat de fizic, după cum arată Davidson. Legile de acest tip pot fi numite *slabe*. Dar legile slabe, nu ar fi suficiente pentru a determina stările mentale, date fiind stările fizice. Așa că pentru ca raționamentul să fie valid, trebuie să presupunem că cel puțin unele dintre legi sunt tari în sensul că ele au forma

$$\forall_x (\emptyset x \rightarrow \psi x)$$

Legile tari ar determina stările psihologice, date fiind stările fizice. Pe de altă parte, nu este clar dacă explicațiile-raționale presupun legi tari.

Totuși, cu condiția ca cel puțin unele dintre legile în cauză să fie tari, raționamentul pe care i l-am atribuit lui Davidson pare să fie valid. Probabil că el este mult mai îndoielnic când premisele (1) și (3) sunt adevărate. Totuși sunt înclinat să le accept. În orice caz, în această interpretare, problema esențială a lui Davidson poate fi pusă mai simplu. după cum urmează: explicațiile-motiv nu pot fi *adevărate în mod obiectiv*, în aceeași măsură în care sunt adevărate explicațiile fizice. Nu știu dacă Hempel ar accepta acest lucru, dar pentru tradiția hermeneutică el ar putea fi potrivit.

Oare Davidson însuși ar accepta această versiune a raționamentului său? Nu pot fi sigur. El spune că nu ar fi de „bun simț” să gândești că atribuiri de atitudini făcute de noi sunt „mai puțin decât obiective” și că o astfel de atribuire, dacă reprezintă cea mai bună atribuire „este obiectivă pe cât poate fi” (1986, p. 205). Interpretarea mea nu neagă că explicația-motiv poate fi „atât de obiectivă cât e posibil să fie”. Sugerez doar că din punctul lui Davidson de vedere, ea este mai puțin corectă în mod obiectiv decât apare posibil în cazul explicațiilor fizice.

Brian McLaughlin menționează o interpretare a raționamentului lui Davidson, foarte asemănătoare cu a mea, dar el susține că nu acest lucru l-ar fi avut în minte Davidson. Totuși, motivul invocat de McLaughlin pentru interpretarea sa nu este convingător. El indică numai că pot exista fapte reale care sunt *relative* la o alegere arbitrară din schema de interpretare. Nu aş nega

acest lucru, dar relativitatea implicată aici reprezintă tocmai ceea ce distinge explicațiile-motiv de explicațiile fizice, în versiunea dată de mine pentru raționamentul lui Davidson.

Dacă versiunea propusă se dovedește acceptabilă și dacă este corectă doctrina lui Quine despre indeterminarea traducerii, pare adevărat să vorbim despre o diferență între „explicație” (în științele fizice) și „interpretare” (în științele umane). În această privință doctrina hermeneutică pare corectă. Totuși, diferența este mai curând ontologică decât logică sau epistemologică și nu implică că modelul-lege-de-acoperire nu ar putea fi aplicat la explicațiile-motiv, sau că ar fi ceva greșit în pozițiile pozitiviste. Doar elementul adițional al subiectivității și interdeterminării prezent în interpretare o deosebește pe aceasta de explicație.

## 9. CONCLUZIE

Am văzut că abordările pozitiviste și hermeneutice ale explicației-motiv pot să nu se contrazică, în fapt, una pe alta. Desigur, adepții fiecărei tradiții resping ideile celeilalte, și pot fi în dezacord și în alte moduri, dar părerea mea este că afirmațiile lor pozitive despre natura explicațiilor-motiv pot fi, toate, acceptate. În privința aceasta, lunga dezbatere-explicație sau înțelegere-poate că este o pierdere de timp.

### *Mulțumiri*

Sunt recunoscător lui Akeel Bilgrami, Dagfinn Follesdal, Hans Mathlein, Paul Needham, Peter Pagin, Philip Pettit, Wlodek Rabinowics, Folk Tersman, Torbjörn Tännsjö și lui Tom Wartenberg, pentru comentarii făcute la prima versiune a acestui

studiu. Mulțumesc de asemenea lui Marx Wartofsky, pentru invitația de a-l prezenta la Baruch Colloquium for Philosophy, Politics and the Social Sciences, în Aprilie 1989.

## NOTE

<sup>1</sup> Vezi, Apel, 1977, p. 167.

<sup>2</sup> Pentru poziția pozitivistă, vezi Hempel, 1965, pp. 239–240. Pentru punctul de vedere hermeneutic, vezi Markovic, 1972, p. 34. Un raționament similar cu al meu este prezentat în Putnam, 1978, pp. 71–75

<sup>3</sup> Vezi Markovici, 1972, p. 31.

<sup>4</sup> Vezi Davidson, 1980, pp. 15–17, 222–223 și 230–233.

<sup>5</sup> Vezi Davidson 1987, pp. 46, unde acesta obiectează justificat la o afirmație mai veche a mea, pe care o retrag.

<sup>6</sup> Vezi Davidson 1980, p. 221–223. Pentru un argument similar, vezi Hempel, 1965, p. 476.

<sup>7</sup> Vezi Davidson 1980, pp. 267, pp. 43.

<sup>8</sup> Vezi Davidson 1985, pp. 245 și Hempel, 1965, pp. 113–116 și 459–462.

<sup>9</sup> Vezi Davidson 1980, pp. 214, 241, 243; 1987, pp. 35 și 46.

<sup>10</sup> Vezi și Davidson 1980, pp. 222–223 și 231. Este greu de înțeles de ce Davidson susține că poziția sa este incompatibilă cu punctul de vedere al lui Føllesdal, conform căruia „teoria noastră despre mental și teoria despre natură sunt, ambele, părți dintr-o teorie mai largă” (1987, p. 44). Putem considera de la sine înțeles că o teorie mai cuprinzătoare a mentalului și a naturii să includă legile psihofizice. Dacă nu este așa, ele reprezintă, în cel mai bun caz, o pereche de teorii separate. Potrivit lui Davidson, legile psihofizice nu pot fi stricte (în sensul lui). Dar el arată și că Føllesdal „are un punct de vedere mai puțin strict despre legi” (1987, p. 44), așa încât teoria cuprinzătoare la care se gândește Føllesdal nu trebuie să conțină legi psiho-fizice *stricte* (în sensul lui Davidson). Dacă este așa, opinia lui Davidson este compatibilă cu cea a lui Føllesdal.

## BIBLIOGRAFIE

APEL, K.O. „Causal explanation, motivational explanation and hermeneutical understanding (remarks on the recent stage of the explanation-understanding controversy), în: G. Ryle (Ed.) *Contemporary Aspects of Philosophy*, (1977) (Londra, Oriel Press) pp. 161–176.



- Davidson, D. (1980) *Essays on Actions and Events* (Oxford University Press).
- Davidson, D. „Judging interpersonal interests“, în: J. Elster, A. Hylland (Eds.) *Foundations of Social Choice Theory* (Cambridge University Press) pp. 195–211.
- Davidson, D. Problems in the explanation of action, în: P. Pettit, R. Sylvan & J. Norman (Eds.) *Metaphysics and Morality; essays in honour of J.J.C. Smart* (1987) pp. 35–49 (Oxford, Basil Blackwell).
- Dray, W.H. *Laws and Explanation in History* (Oxford University Press) (1957).
- Dray, W.H. „The historical explanation of actions reconsidered“, în S. Hook (Ed.) *Philosophy and History. A Symposium*, (1963) pp. 105–135 (New York University Press).
- Follesdal, D. „Causation and explanation a problem in Davidson's view on action and mind“, în: F. LePore & B. McLaughlin (Eds.) *Actions and Events: perspectives on the philosophy of Donald Davidson*, (1985), pp. 311–323 (Oxford, Basil Blackwell).
- Hempel, C.G. „Rational action“, *Proceedings and Addresses of the American Philosophical Association*, (1962) 35, pp. 5–23
- Hempel, C.G. *Aspects of Scientific Explanation* (New York, Free Press). (1965)
- Kim, J. „Psychophysical laws“, în: F. LePore & B. McLaughlin (Eds.) *Actions and Events*, (1985) pp. 369–386 (Oxford, Basil Blackwell).
- Markovic, M. „The problem of reification and the Verstehen Erklären controversy“, în *Acta Sociologica*, (1972) 15, pp. 27–38.
- McLaughlin, B.P. „Anomalous monism and the irreducibility of the mental“, în: F. LePore & B. McLaughlin (Eds.) *Actions and Events*, (1985) pp. 331–368 (Oxford, Basil Blackwell).
- Mill, J.S. (1862) *A System of Logic*.
- Putnam, H. *Meaning and the Moral Sciences* (London, Routledge & Kegan Paul) (1973).

Traducere Maria VORNICU și Angela BOTEZ

# FILOSOFIA ANALITICĂ ȘI TEORIA PSIHANALITICĂ \*

Graham J. LOCK

(Universitatea din Oxford)

Articolul de față tratează teme de intersecție, cum ar fi relația între filosofie și teoria psihanalitică, critica lui Wittgenstein asupra lui Freud, abordarea filosofică post-wittgensteiniană a concepției lui Freud, filosofia post-wittgensteiniană a limbajului în lumea anglo-saxonă și relația ei cu opera lui Frege; paralela între filosofia anglofonă a limbajului, pe de o parte, și anumite tendințe ale gândirii franceze în domeniu, pe de altă parte.

Mulți filosofi din Anglia și America sunt interesați de psihanaliză. Dacă îl considerăm pe Ludwig Wittgenstein ca filosof al lumii anglofone <sup>1</sup> trebuie să spunem că dezbaterea în jurul doctrinei lui Freud ocupă un loc semnificativ în lucrările filosofice de acolo. Ca și Freud, Wittgenstein a fost vienez. El era familiarizat mai mult sau mai puțin cu lucrările lui Freud, sora sa Margarete a fost pacienta și prietena lui Freud. Wittgenstein a scris puțin despre Freud, dar s-a referit frecvent la lucrările acestuia în cursuri și conversații cu studenții și prietenii.

În *Lecturi și conversații asupra esteticii* și în *Psihologie și credință religioasă* el vorbește despre „ceva ce Freud susține și mie mi se pare total greșit“. Problema se referă la natura explicației psihanalitice. Freud, spune Wittgenstein, sugerează că uneori comportamentul poate fi explicat prin descoperirea, dedesubtul motivului conștient, a motivului adevărat și inconștient. Dar, pe baza cărui criteriu se determină că cel de al doilea motiv și nu primul este cel eficient? Ar fi ca și cum Freud a

---

<sup>1</sup> Din articolul „Analytic Philosophy, Psychoanalytic Theory and Formalism“, *Revue de synthèse*, nr. 2/1987.

făcut o descoperire și, în același timp, tot el a introdus o metodă pentru aplicarea ei la cazuri individuale. Wittgenstein consideră că nu există o descoperire la Freud, ci mai degrabă un proces de persuadare – a pacientului de către analist – care de multe ori funcționează: pacientul fiind convins că în fapt acționează din motive diferite de ceea ce credea el înainte că stă în spatele acțiunii sale. În acest sens, psihanaliza oferă pacientului o nouă și diferită perspectivă asupra lumii, nu cu mult mai adevărată decât cea dinainte, dar cu valoare potențială morală sau terapeutică. Acest argument este tipic pentru o anumită linie a criticii aduse interpretării „științifice” dată de Freud însuși operei sale, dar este, în același timp, bazată pe o cunoaștere superficială a textelor lui Freud. Wittgenstein pare să creadă, judecând după comentariile sale din 1938, că teoria lui Freud este un pansexualism și, chiar că ea ar susține faptul că anxietatea este întotdeauna explicabilă ca urmare a unei traume din naștere. Dar aceasta reprezintă mai degrabă concepția „eretică” a lui Otto Rank.

Wittgenstein pare a fi fascinat de întreaga chestiune a „semnificației” viselor și de lucrările lui Freud în acest domeniu. Imaginile de vis, notează el, par a avea „o anumită asemănare cu semnele limbajului”. Dar, dacă ele sunt într-adevăr simboluri, cum susține Freud, și pot fi traduse în limbajul vorbirii obișnuite, „atunci traducerea trebuie să fie posibilă în ambele direcții”. Până acum, după cum recunoaște, chiar Freud, nimeni și nici analistul în particular nu este capabil să traducă „gânduri obișnuite” în „limbajul viselor”. Ceea ce se poate face reprezintă traducerea viselor post facto, în limbaj obișnuit. Deci, nu există în realitate un limbaj al viselor.

Dar critica merge mai departe. Freud caută nu numai traducerea sau interpretarea visului, a unei remarci sau a unui simptom, ci „interpretarea lor corectă”. De ce ar crede el că există doar o singură interpretare corectă? La această întrebare ar trebui să se răspundă cu o altă întrebare: „Îmi cereți să cred că există ceva care se întâmplă fără o cauză?” a identifica o cauză înseamnă a da adevăratul răspuns la întrebarea relevantă „de ce”? Dar acesta, obiectează Wittgenstein, ar fi un răspuns irelevant, pentru că noi nu trebuie să căutăm cauza în „sensul unui eveniment trecut” conectat într-un mod potrivit cu visul, remarcă, sau simptomul. Tehnicile psihanalitice (cum ar fi asociația liberă) ne dau posibilitatea să descoperim anumite lucruri despre noi înșine dar nu „ne explică de ce apare un anumit vis (etc)”.

Se pare că Wittgenstein a recitat *Interpretarea viselor* în 1946, și a ajuns la câteva concluzii cu privire la doctrina lui Freud. Una ar fi că Freud a propus, fără a admite și fără a fi conștient de acest lucru, „un nou mit“ care este de fapt adevăratul statut al psihanalizei. Trăgând o astfel de concluzie nu înseamnă să condamni teoria lui Freud, pentru că mitul are funcții sociale utile. Se ridică însă din nou problema statutului științific al teoriei. În jurul întrebării: Este psihanaliza o știință și dacă da, în ce sens? se poartă de mulți ani dezbateri în cercurile filosofice anglosaxone – dezbateri care au și astăzi o considerabilă răspândire.

În 1948, Wittgenstein se reîntoarce la chestiunea interpretării viselor. Ceea ce intrigă cu privire la un vis, scrie el, nu este „conexiunea lui cauzală cu evenimentele vieții mele, ci mai degrabă impresia pe care mi-o dă că ar fi un fragment al unei povestiri... care rămâne în rest obscură. Și această poveste visată are „un șarm al ei propriu“<sup>2</sup>. În *Investigații filosofice*<sup>3</sup> își îndreaptă din nou atenția spre aplicabilitatea conceptului de „cauză“ la anumite feluri de explicație a comportamentului, – dar acum comentariile sale sunt mai favorabile metodei lui Freud.

Să presupunem, spune el, că am ieșit să mă plimb afară din oraș și după ce am hoinărit o vreme îmi imaginez că orașul este undeva la dreapta mea. „Deși nu am nici un motiv rațional să cred asta, mi se pare totuși că există anumite cauze psihologice pentru aceasta: în particular anumite asociații și amintiri... Va trebui să încerc să „descopăr psihanalitic cauza convingerii mele nefondate“.

Mult mai înainte, în „*Cursurile de la Cambridge*“ din 1932–1935, Wittgenstein afirmase că a identificat la Freud „O confuzie... între cauză și rațiune: A fi clar de ce râzi nu înseamnă că este clară și cauza... Noi putem găsi cauze ale râsului“; care pot fi descoperite experimental, pe când calea psihanalitică a descoperirii motivului pentru care cineva râde nu implică experimentul, ci doar procedura prin care acea persoană ajunge să fie în final de acord cu rațiunea pentru care a râs. Iar acest tip de procedură nu este știință, este pur și simplu, o (nouă) „modalitate de reprezentare“ a lucrurilor. Și G.E. Moore, susținea Wittgenstein, a apărut distincția între cauză și rațiune în contextul discutării teoriei freudiene.

În *Carnetul Albastru* din 1933–1934, Wittgenstein caracterizează folosirea de către Freud a noțiunii de „idei inconștiente“ ca

pe o inovație mai ales terminologică. Cineva ar putea fi abătut de la calea cea bună prin imaginile și analogiile sugerate de noua terminologie, spune el. Prin comparație, să presupunem că cineva a introdus fraza „durere de dinți inconștientă“, frază care poate avea un sens. Reacția noastră poate fi însă de negare a oricărei posibilități a unui astfel de lucru, dar putem fi și copleșiți de prodigiozitatea descoperirii. De fapt, este vorba de o simplă alegere a modului de descriere cel mai folositor și avantajos. Un psihanalist poate, în mod similar, greși deplin, dacă presupune că a descoperit „gânduri conștiente care au fost (de fapt) inconștiente sau ceva confuz de acest fel.

Pe de o parte, Wittgenstein pare a fi sceptic cu privire la tezele științifice ale psihanalizei, considerându-le periculoase în practica clinică. El îi spunea studentului său Norman Malcolm, în 1945, că psihanaliza a produs o enormă stricăciune și foarte puțin bine. Pe de altă parte, el se consideră totuși un discipol al lui Freud, oferind de fapt trei explicații pentru popularitatea teoriei acestuia. Prima: în existența umană se manifestă o fascinație a „nefirescului“<sup>4</sup> care este satisfăcută de psihanaliză. A doua: oamenii au o tendință iconoclastă, flatată de practica și teoria care îndrăznește să vorbească despre tabuuri sexuale. A treia se referă la faptul că atracția psihanalizei constă, în parte, și în mod paradoxal, în caracterul său „respingător“. O explicație de același tip poate fi găsită încă la Platon, în *Republica* (IV, 439–44). Se menționează o poveste privind pe Leontin fiul lui Aglaion, care împiedicându-se de niște cadavre umane, a simțit dorința să se întoarcă să se uite la ele, în același timp fiind oprit de dezgust. În cele din urmă dorința a învins și el a venit la cadavre și le-a deschis ochii. Platon și Wittgenstein acceptă amândoi existența fenomenului pe care unii dintre filosofi de limbă engleză par a-l nega. Mary Midgley, de pildă, în lucrarea sa *Wickedness* susține că nu este plauzibil să presupui că moartea ar putea fi un obiect al dorinței și, mai general, că numai motivele sănătoase pot fi inteligibile<sup>5</sup>.

Anthony Kenny notează<sup>6</sup> că există, dincolo de atracția lui Wittgenstein spre psihanaliză, o asemănare între metoda psihanalitică și practica filosofică a lui Wittgenstein. Filosofia, a crezut Wittgenstein, este folosită numai oamenilor care sunt într-un anume fel bolnavi, aceia care au în interior o filosofie rea. Dar

aceasta este adevărat pentru noi toți. Filosofia poate fi deci denumită „terapeutică“. Wittgenstein însuși o compară cu medicina.

Morris Lazerowitz încearcă să dezvolte această idee într-un număr de lucrări, în special în *Structura metafizicii* (1955), *Studii de metafilosofie* (1964), *Filosofie și iluzie* (1968), și *Limbajul filosofiei: Freud și Wittgenstein* (1977).

Filosofia, susține Lazerowitz, este ca un sanctuar unde ceva ce sună ca un discurs aberant, poate fi acceptat ca un discurs rațional. Wittgenstein susține aceasta în lucrarea sa *Asupra certitudinii* (On Certainty). El ne spune povestea unui filosof care stând în parc arată un pom și afirmă: „Eu știu că acesta este un pom“ iar către un trecător care îl aude: „Nu sunt un tip bolnav, doar făceam filosofie“. O filosofie rea, avea să observe Wittgenstein în altă parte, se găsește sub formă de mitologie în limbajul de zi cu zi<sup>7</sup>. Dar aceasta nu înseamnă că scopul filosofiei trebuie să fie crearea unui *limbaj ideal*. Aici atingem un punct de o oarecare importanță. Interesul lui Wittgenstein pentru psihanaliză pare în parte inspirat din dorința de a „trata“ oamenii de boala filosofiei rele. Un alt exemplu de filosofie rea ar fi căutarea limbajului ideal – un proiect foarte vechi. Printre filosofi din epoca modernă care l-au susținut se numără și Leibniz. Astăzi cercetările au luat un nou avânt datorită orientării calculului logic, care – se speră – va putea elimina în final defecțiunile logice inerente limbajului natural. Acest proiect a fost respins de Wittgenstein în mai multe lucrări<sup>8</sup>.

Oare Freud nu ne îndemna împotriva aceleiași tentații, în lucrarea sa *Noi lecturi introductive*, când scrie: „Filosofia nu este opusă științei, ea se comportă ca și știința și lucrează în parte prin aceleași metode, dar se depărtează de științe prin iluzia de a fi capabilă să prezinte o imagine coerentă a universului fără *lacune*. Ea merge aiurea prin supraestimarea valorii epistemologiei a operațiunilor noastre logice“<sup>9</sup>. Este dificil de a înțelege exact ce intenționează să exprime Freud cu aceste cuvinte. Ele sunt probabil congruente cu poziția lui Wittgenstein și includ o opoziție hotărâtă față de ceea ce se cheamă uneori „perspectivă calculatorie“ a limbajului și față de repercursiunile ei asupra filosofiei și a altor întreprinderi intelectuale. Astăzi, această poziție are consecințe importante asupra problemelor teoretice, inclusiv aceea a naturii explicației psihanalitice...

...Să ne întoarcem în Franța. Cum am arătat deja<sup>\*</sup>, și aici Frege a reînceput să joace un rol în istoria intelectuală recentă, fapt care are legătură cu încercările lui Lacan de fundamentare filosofică a „întoarcerii lui Freud“ E. Roudinesco remarcă faptul că Lacan a avut dorința de a găsi o cale pentru finalizarea doctrinei psihanalitice. Cercetarea sa a luat prin 1969 forma studiului intensiv al *Tratatului* lui Wittgenstein, dar faptul dacă el a înțeles sau nu sensul acestei lucrări a rămas neclar. În orice caz, Lacan și colaboratorii săi intimi au fost inspirați de Leibniz și Frege în încercarea de a produce un adevăr universal, o teorie științifică (formalizată) a Inconștientului. În Franța, ca și în Anglia și SUA interesul reînnoit pentru acești gânditori se leagă de dezvoltarea teoriei limbajului. Roudinesco a rezumat comparativ situația la mijlocul anilor 1966: „Prin contrast cu situația din țările anglo-saxone, unde moștenirea fregeeană domină scena epistemologică, Franța epocii acesteia este în urmă pe terenul logicii. În 1965, lucrările lui Frege și ale succesorilor săi rămân prost cunoscute de o intelectualitate marcată de o altă tradiție în materie de istorie a științelor. Totuși odată cu explozia structuralistă se produce și o nouă interogare asupra logicii, în măsura în care studiul structurilor lingvistice tinde spre cercetarea fenomenelor generale ale gândirii și limbajului<sup>10</sup>. În această situație Lacan, influențat de lecturile sale din Saussure și Jakobson a decis să-și intensifice studiile de teorie socială și lingvistică. Urmându-l pe Jakobson el a respins abordarea chomskiană. A introdus apoi o anumită interpretare, diferită însă, pentru a-și legitima doctrina finalistă. El pare să fi gândit ca Wittgenstein din perioada *Tratatului*, pe de o parte și ca Frege, pe de alta. Cei doi par a fi fost surse complementare pentru proiectul lacanian dar de fapt, au rămas departe de el<sup>11</sup>. În seminarul său din 1969–1970, „L'Envers de la psychanalyse“, au fost puse fundamentele pentru formalizarea teoriei psihanalitice. Un prim pas, probabil inspirat de Leibniz și a sa *mathesis universalis*, este căutarea unor matheme psihanalitice ceva asemănător unui set de ecuații sau formule care să permită formalizarea teoriei psihanalitice (și prin aceasta, trecerea dincolo de tratamentul analitic, în universalitate, de exemplu).

---

<sup>\*</sup> În traducerea de față a fost redusă partea despre logica lui Frege și ca atare și notele respective.

Lacan a negat întotdeauna noțiunea de „ego“ cartezian. Noțiunea de „mathemă“ vine să umple golul lăsat de aceasta. Ea a introdus, cel puțin în principiu, posibilitatea unei transmisii non-conflictuale a unei doctrine. Astfel, spiritul lui Leibniz a reușit, gradat, să obțină dominația asupra școlii lui Lacan. Închizând ușa cartezianismului Lacan a permis unei versiuni a formalismului logic să intre pe ușa din dos. Cuvintele lui F. Gadet și M. Pécheux sunt relevante cu privire la sensul sintagmei „critica limbajului“. Ei se referă la un demers destinat a controla limbajul în scopul de a-l elibera de elemente iraționale pentru a-l reforma și transforma în limbaj conform cu principiile rațiunii. Această tendință bazată pe filosofia logică care reînnoiește legătura cu concepția leibniziană caută un limbaj ideal cu caracteristici universale și este în fapt o versiune a neopozitivismului<sup>12</sup>. Roudinesco merge mai departe și afirmă că logicismul „mathemelor“ este maniera în care neo-pozitivismul (dominant de cealaltă parte a Atlanticului) s-a infiltrat în Franța în discursul psihanalitic<sup>13</sup>.

Există deci o paralelă între nașterea, în lumea intelectuală de limbă engleză și în cea franceză, a neoleibnizianismului și a neofregeianismului. Orientări de tipul celei lacaniene sunt supuse criticii, ca de exemplu cea specifică lui Wittgenstein. Vreau să subliniez pe scurt motivele pentru care cred eu că această doctrină este vulnerabilă.

Lacan redescrive procese fizice identificate de Freud: ca de pildă, de cathesis, cathesis liber și re-cathesis care înseamnă că o cantitate de energie psihică este detașată de la o idee dată și reatașată la o nouă idee (de obicei se vorbește de acest proces în cazul fobiilor), deși Lacan preferă să vorbească de procesul metaforic. După părerea lui Lacan ceea ce se întâmplă face ca un semnificator să fie înlocuit cu altul, primul fiind trimis în inconștient. Aici își pierde funcția originală de semnificare și devine o „image neverbalizată“, ceea ce Freud numește „idee concretă“. Nu este vorba de faptul că inconștientul ar impune o nouă structură. La Lacan inconștientul este structurat ca un limbaj. Structura sa sintactică nu are istorie sau caracter cultural specific, fiind universal prezentă la toate societățile și persoanele.

Tot ceea ce este psihanalitic interesant despre comportamentul oamenilor va putea fi exprimat în termenii „limbajului inconștientului“ și a legilor lui. Aici încep însă dificultățile. Pentru ce tip



de explicație se va opta, pentru una de tip *causal*? Pe baza celor spuse mai sus, legile lingvistice ale inconștientului nu pot să devină reguli ale limbajului, pe care o persoană să le urmeze sau nu. Persoana nu poate nici să le aplice, nici să le violeze, pentru că îi sunt necunoscute, chiar dacă îi determină comportamentul. Vorbind totuși de legi universale, pe care nimeni însă nu le poate învăța, Lacan crede că ar reuși să le formuleze în limbajul artificial formalizat al *mathemelor*. Limbajul formal al inconștientului nu poate însă avea legi, dacă ar fi avut ar trebui să existe un agent care să le aplice sau nu. Dacă el ar funcționa în acord cu legile cauzale ar însemna că posedă un mecanism causal pe care teoreticianul în psihanaliză ar putea încerca să-l descopere.

Abordarea lui Lacan invocă o concepție comparabilă cu filosofii limbajului, devenite populare în spațiul de limbă engleză în ultima vreme, dar care au fost aspru criticate (în particular de G.P. Baker and P.M.S. Hacker). Unul din punctele centrale atacate de critici<sup>14</sup> se referă la faptul că este un nonsens să vorbești de un limbaj care funcționează în acord cu reguli despre care nimeni nu este conștient. Nu poți urma reguli pe care nu le înțelegi. Un savant nu poate descoperi că cineva a urmat o lege în mod inconștient, dacă legea îi era necunoscută. Regulile nu sunt ipoteze explicative. Există teorii dar numai pentru a explica comportamentul guvernat de reguli. Alternativa ar consta în abandonarea ideii că „limbajul inconștientului funcționează în acord cu un mecanism causal“, dar atunci acesta nu mai poate fi numit limbaj. Putem, de asemenea, așa cum sugerează Wittgenstein, să extindem conceptul de limbaj, astfel încât să acopere lucruri noi, cum ar fi „limbajul computerului“ și „limbajul naturii“. Dar, atunci, sensul conceptului de limbaj se schimbă chiar în aspectul care ne face să înțelegem cum funcționează el, astfel încât poate fi învățat printr-un proces de încercări și erori și prin aplicarea sa corectă sau incorectă de către cel care-l practică.

Ceea ce Lacan pare a încerca este tocmai ceea ce acest argument exclude. Concepția sa asupra inconștientului pare să fie asupra a ceva care funcționează „ca un limbaj“ pe baze de reguli, dar care, în fapt, operează ca un mecanism causal. Un astfel de mecanism nu se dovedește potrivit pentru explicarea comportamentului intențional și finalist, conștient sau inconștient. Aceasta este concluzia pe care poate să o tragă cineva care urmează linia generală a lui Wittgenstein „cel târziu“.

Să ne întoarcem, în încheiere, la Wittgenstein ca autor al *Investigațiilor filosofice*. Scopul său aici a fost să se opună concepției calculatorii asupra limbajului. El încearcă nu numai să arate greșelile acestei viziuni, dar să și explice popularitatea ei. Cum arată Brian McGuinness, într-un comentariu asupra lui Wittgenstein, oamenii cred că dacă dau nume lucrurilor le și stăpânesc. Ei obțin iluzia grijii unui „înger al cunoașterii” și le place. „Trebuie însă rezistat acestui fals paradis, eliberându-ne de limbajul indus de imaginea mitologică în primele noastre reflecții asupra lumii”<sup>15</sup>. Prin astfel de afirmații, scrie McGuinness, Wittgenstein se declară discipol al lui Freud în ideea că omul se lasă flutat de sensuri aparente și captivat de mitologie<sup>16</sup>. Dacă Wittgenstein este un discipol al lui Freud, care nu a făcut altceva decât să-i facă obiecții, Lacan este un discipol care a cerut întoarcerea la ortodoxia freudistă. Problema—care dintre ei a fost, în spirit, mai aproape de Freud—rămâne deschisă. Putem spune că raportat la Lacan-ul anilor 1970 Wittgenstein (care a murit în 1951) poate fi numit anti-Lacan „avant la lettre”.

## NOTE

<sup>1</sup> Născut și crescut în Austria, a lucrat exclusiv în Anglia.

Cf. A. Janik, St. Toulmin, *Wittgenstein's Vienna*, New York, Simon și Schuster, 1973.

<sup>2</sup> *In Culture and Value*, Oxford, Blackwell, 1980, pp. 68 – 69.

<sup>3</sup> *Philosophical Investigations*, Oxford, Blackwell, 1953 p. 215.

<sup>4</sup> Vezi *Lectures and Conversations*, Oxford, Blackwell, 1966 p. 25,

Cf. cu articolul lui Freud „Das unheimliche” tradus ca „Un canny”, în Ediția standard, Londra, Hogarth, 1974.

<sup>5</sup> *Wickedness: A Philosophical Essay*, Londra, Routledge 1984.

<sup>6</sup> „Wittgenstein on the Nature of Philosophy” în Brian McGuinness (ed) *Wittgenstein and his Times*, Oxford Blackwell, 1982.

<sup>7</sup> *In Manuscript 213*, p. 115 (Pentru informații asupra manuscriselor lui Wittgenstein și numărul lor, vezi G.H. von Wright „The Wittgenstein Papers” în *Wittgenstein*, Oxford, Blackwell, 1982).

<sup>8</sup> Vezi, de exemplu, *Blue and Brown Books*, Oxford Blackwell, 1958, p. 28; *Philosophical Investigations* secția 132.

<sup>9</sup> Conferința XXXV, în *Standard Editions*, op. cit. 4, 1964, vol. XXII.

<sup>10</sup> E. Roudinesco, *Histoire de la psychanalyse en France*, 2, 1925 – 1985, Paris, Seuil, 1986, p. 401.

<sup>11</sup> Cf. cu G.P.Baker, P.M.S.Hacker în, *Language, Sense and Nonsense*, Oxford, Blackwell 1984, pp. 373 – 374.

„Wittgenstein a criticat concepția adevărului logic prezentă la Frege și Russell (...) Acești filosofi care au fost convertiți la concepția semantică modernă a validității sunt direct influențați de *Tratat*, nu de Frege“. La pag. 380: Întreaga concepție a condițiilor de adevăr (...) a stârnit un mare interes când a fost publicată de Wittgenstein, dar ea este foarte departe de concepția ce se poate identifica în lucrările lui Frege. Conceptul (de condiție a adevărului, explicat de Wittgenstein în *Tratat*) nu coincide cu concepția despre sens a lui Frege ș.a.m.d.

<sup>12</sup> Françoise Gadet, Michel Pécheux, *La langue introuvable*, Paris, Maspero, 1981, p. 114.

<sup>13</sup> E. Roudinesco, *Pour une politique de la psychanalyse*, Paris, Maspero, 1977, p. 113.

<sup>14</sup> În particular, de către G.P.Baker, P.M.S.Hacker, *op. cit.*

<sup>15</sup> Cf. cu *Freud and Wittgenstein*, în B McGuinness (ed) *op. cit.* pp. 42 – 43. Vezi, de asemenea, Jacques Bouveresse „Wittgenstein face á la psychanalyse“ *Austriaca*. (Univ. de Haute-Normandie) 21 nov. 1985, pentru clarificări ale unor probleme ridicate de acest articol 37.

<sup>16</sup> Acest articol se bazează în parte pe seminarul ținut la Universitatea din Oxford, Subfacultatea de Filosofie și la Colegiul internațional de filosofie de la Paris, instituții cărora le mulțumesc pentru ospitalitate.

Traducere Angela BOTEZ

# CONȘTIINȚĂ DE SINE ȘI CUNOAȘTERE MORALĂ

Liubomira MIROȘ

(Institutul de filosofie, București)

**Conștiința de sine morală** reprezintă un stadiu superior, calitativ deosebit, al dezvoltării conștiinței de sine.

Noțiunea conștiinței de sine a căpătat sensuri diferite, atât în dezvoltarea științei cât și în istoria gândirii filosofice.

Încă în 1799, medicul și filosoful francez F.X. Bichat distinge două tipuri de conștiință: 1) o conștiință exprimabilă în gândire precisă și limbaj, ca o conștientizare a situației individului în lume; 2) o conștiință vagă, ca afirmare originară a unei spontaneități biologice constitutive. În cadrul neurobiologiei, cercetările ulterioare urmăresc explicarea interrelației dintre cele două tipuri de conștiință, în strânsă conexiune cu caracteristicile celor două sisteme nervoase – central și neurovegetativ – și cu raporturile dintre ele. Tendința manifestată în știință se caracterizează, în general, prin încercarea de a privi fenomenul conștiinței de sine în *ansamblul* organismului, în perspectiva unității persoanei. În secolul XIX, cercetările empirice privind conștiința de sine s-au efectuat în cadrul psihologiei personalității. Cercetările descriptive sunt, în această perioadă, slab legate de teorie. O dezvoltare accelerată a cercetărilor are loc în secolul nostru. Între anii 1940–1970, au apărut aproximativ 2 000 de lucrări de psihologie dedicate cercetării problematicei „eului”. Obiectul cercetărilor îl constituie, în special, structura și componentele imaginii „eului”, procesele și operațiile psihice prin care individul se cunoaște pe sine, se apreciază și se controlează, comportamentul său: funcțiile psihologice ale conștiinței de sine (gradul de adecvare a imaginii eu și a aprecierilor izolate, ca și rolul pe care acestea îl dețin în autoreglarea comportamentului individual).

Prin relația stabilită de Hegel între *datoria pură* și *sensibilitate*, se delimitează *cadru* global al structurii conștiinței de sine. Urmează elementele și nivelele structurii, care reprezintă momente integrate în procesul constituirii sale, drept cunoaștere de sine și reflectare de sine *morală*. Este vorba, în linii mari, de nivelele: *biologic* – instincte, impulsuri și trebuințe vitale; *biopsihologic* – schema corporală sau conștiința biologică globală; *psihologic* – eu și conștiința de sine ca reflectare asupra eului; *psihomoral* – imaginea eu. Ultimul nivel le integrează pe cele anterioare, reprezentând nivelul care sintetizează condițiile de posibilitate pentru trecerea la conștiința de sine morală. Calitatea conștiinței de sine morale depinde de natura, funcția și modalitatea constituirii și ființării imaginii eu, sau imaginii de sine. Aceasta constituie condiția sa psihologică fundamentală.

Eul prezintă trei determinări principale: *identitatea*, care se păstrează și în condițiile schimbării mediului; *subiectivitatea*, care cuprinde mecanismele de reglementare și autoreglementare ale comportamentului și motivațiilor interioare; „*imaginea eu*“, care corectează comportamentul personalității și reprezintă gradul de adecvare a autoaprecierii<sup>1</sup>.

Asigurarea adecvării autoaprecierilor prezintă dificultăți datorate însuși specificului acestora. „Autoaprecierile – consideră I. S. Kon – sunt scheme cognitive proprii, care generalizează experiența trecută și organizează, structurează noua informație privitoare la determinarea dată a eului“<sup>2</sup>. Procesul de structurare și organizare nu se reduce însă numai la un proces de cunoaștere, ci presupune valorizare, cu determinările sale subiective și valorice specifice. Cu alte cuvinte, conținutul și modul în care se structurează schemele cognitive, în care sunt organizate și ierarhizate elementele acestora, presupun cu necesitate *selecția* informației, în funcție de emoții, aspirații, valori, model de personalitate, pretenții și succese, eșecuri și renunțări, comparația cu alții etc. În această intervenție a subiectivității și afectivității, constă în esență, posibilitatea readecvării autoaprecierii la conținutul obiectiv supus cunoașterii și autocunoașterii.

În procesul autoaprecierii se realizează, astfel, unitatea dintre *identitate*, dintre ceea ce este persoana și *schimbarea*, sau ceea ce

devine sau vrea să devină persoana. Un aspect important al procesului îl reprezintă măsura în care sunt conștientizate însușirile individului, câte și care anume dintre ele, ca și gradul conștientizării lor. Un rol și mai important în formarea imaginii-eu, îl deține conștientizarea relației dintre felul în care individul încearcă să apară în proprii săi ochi și în ochii celorlalți. pe de o parte, și adevăratul său fel de a fi, pe de altă parte. Dificultatea constă în posibilitatea, în general redusă, a individului, de a adecva ceea ce posedă ca dat natural, moștenit sau asimilat anterior și modelat în experiența de viață, pe de o parte, la ceea ce dorește să devină, la modelul ales de personalitate, pe de altă parte. Pe plan moral, acest lucru se exprimă în jocul dintre sinceritate și luciditate, pe de o parte și duplicitate și autoamăgire, pe de altă parte, între exprimarea a ceea ce este individul și încercarea de apărare așa cum ar dori să fie sau cum ar fi bine și util să apară în ochii celorlalți. Abilitatea de a stăpâni regulile jocului, de a stabili echilibrul dintre constanță și fluiditate, de a sesiza măsura exactă sau măcar aproximativă în care autosugestia deține în același timp și o funcție de automodelare, de a marca limita dincolo de care dorința exprimată în comportament, de a părea altfel, ca mijloc de a deveni în mod obiectiv mai bun, se transformă în ipocrizie și autoînșelare, reprezintă un aspect esențial al funcției de control a autodezvoltării, exercitat de conștiința de sine morală.

Ținându-se seama de caracterul eterogen și complex al elementelor și factorilor care contribuie la formarea autoaprecierilor, se poate formula o problemă cheie de a cărei rezolvare depinde cunoașterea lor și anume *definirea condițiilor în care poate fi realizată adecvarea dintre momentele cognitive și cele valorizatoare în procesul formării unei reprezentări adecvate, adevărate, despre sine.*

Modul în care individul „se vede“ pe sine, depinde de: 1. ce anume și cât cunoaște și alege din propria sa individualitate; 2. ce relație stabilește între sine și lume, societate, familie, colegi de școală, muncă etc.; 3. cum valorizează și după ce criterii, datele obținute în raport cu propriul său eu.

În ceea ce privește datele supuse valorizării, acestea provin din informațiile interioare și cele exterioare. Primele privesc starea organismului în unitatea structurilor sale, care determină moti-

vația interioară a autoaprecierii. Aceasta poate fi conștientă și inconștientă. Cunoașterea de sine poate fi, la rândul său, adevărată sau falsă, falsitatea fiind determinată fie de o cunoaștere incompletă, fie de una greșită sau denaturată. Caracterul denaturat provine nu atât și nu numai din *cunoașterea* denaturată, greșită a componentelor biopsihologice și a interrelației dintre ele, ci și de o modalitate inadecvată a *relației dintre cunoașterea și valorizarea lor*. Cu alte cuvinte, în procesul complex cunoaștere-valorizare, starea, locul și funcția unui element sau al altuia al structurii individualității, pot fi fals cunoscute și apreciate, datorită *criteriilor după care el este supus judecății*: modelul omului, concepția despre om, idealul moral și sensul vieții etc. Ideea este surprinsă, în esența sa, de M. Merleau Ponty – într-un alt context problematic, care, pornind de la constatarea că omul este ființa care nu se mulțumește să coincidă cu sine ca un lucru, ci se reprezintă pe sine, se vede, se imaginează, consideră că *orice schimbare în reprezentarea despre om constituie o schimbare a omului*<sup>3</sup>. Ideea poate fi interpretată într-un dublu sens: pe de o parte, schimbarea omului și necesitățile practicii sociale și implicit morale duc la noi solicitări față de om și de aici la schimbarea concepției despre om, teoriei revenindu-i sarcina de a surprinde și explica modificările obiective survenite și consecințele lor pe diferitele planuri ale existenței umane și, pe de altă parte, orice nouă concepție influențează imaginea despre sine a omului, determinând o altă atitudine față de sine, un alt comportament și alte relații umane, stimulând și încurajând dezvoltarea unor trebuințe și aspirații noi, care aduc după sine modificări în structura omului ca și în relațiile și mecanismele care o determină.

În contextul problematic care ne interesează, un loc deosebit îl dețin concepțiile privitoare la natura corporală și spirituală a omului, la relația dintre ele și valoarea lor în definirea personalității umane. Merită subliniate și concepțiile mai puțin cunoscute (în literatura de specialitate filosofică și istoria artelor acordându-se o atenție deosebită momentului Antichității și Renașterii), și anume cele ale apologeților creștinismului timpuriu, ca primi pași spre o valorizare integrală a omului. Dintre ele se remarcă cu deosebire cele ale lui Tertulian și Lactanțiu. Aceștia combat atitu-

dinea agresiv negativă față de trup și trebuințele sale – ca o reacție la excesele manifestate în această direcție în antichitatea târzie – proprie în special dualiștilor orientali, gnosticilor și unor gânditori ai creștinismului timpuriu. Tertulian susține ideea dependenței sufletului, a forței active și eficienței acestuia, de bunăstarea trupului, ca și cea a conlucrării și colaborării lor în asigurarea calității vieții umane. Trupul nu este considerat, în această viziune, un simplu mijloc și o condiție a sufletului, ci una din cele două părți constitutive ale omului, având, deci, aceeași valoare cu sufletul. De aici, ideea etică prezentă în concepția lui, după care nu numai în trup, ci și în suflet rezidă cauza păcatului. Idee cu consecințe deosebit de importante pentru conștiința de sine și conștiința de sine morală, așa cum se va vedea în continuare, cu deosebire prin sublinierea contrastului prezent în structura „imaginii“.

La argumentele de ordin ontologic și etic, un alt apologet, Lactanțiu adaugă temeuri estetice aprecierii rolului pozitiv al trupului în îmbogățirea și perfecționarea creației. Elogiul trupului prin descrierea frumuseții și perfecțiunii fiecăreia din părțile sale componente se înscrie, în opera lui Lactanțiu, ca o încercare de a valoriza calitățile corporale, alături de cele spirituale, ca elemente ale valorii estetice, ca determinare esențială a ordinii universale.

Ideile menționate sumar contribuie la delimitarea unor caracteristici specifice proprii sensului conștiinței de sine și conștiinței de sine morale. În primul rând, în aprecierea modalității de raportare și integrare a individualității în armonia universului intră, ca parte cu posibilități și consecințe egale – deși diferite – și trupul, alături de spirit. Omul participă activ la perfecțiunea și perfecționarea lumii nu numai prin spiritul și valorile acestuia, ci și cu cele materiale, legate de natura sa corporală.

De aici o anumită înțelegere a relației dintre datorie și sensibilitate ca determinare a conștiinței de sine morale. În lumina ei, imoralitatea, căderea omului în păcat, nu mai pot fi atribuite manifestării sensibilității în forma sa carnală, simplei ei prezențe – ca în concepția unor apologeți și în atitudinea practică proprie ascezei –, ci considerate ca o necunoaștere sau nerespectare a necesității de a stabili o relație între datorie și sensibilitate, care



să țină seama de specificul și cerințele concrete, individuale ale fiecăreia dintre ele, de concordanțe sau discordanța lor. Condiția împlinirii datoriei pure, nu rezidă, în acest caz, în rușinarea de cerințele trupului și reprimarea lor, ci în înnobilarea și folosirea acestora, prin valorizarea contribuției lor reale la dezvoltarea armonioasă și deplină a omului. Datoriile concrete, care specifică datoria pură, printre care datoriile față de sine ocupă un loc important, vor număra astfel la rândul lor și datoriile față de trup, față de componentele, trebuințele și capacitățile acestuia, manifestându-se în acțiunea pentru cunoașterea, valorizarea și satisfacerea acestora.

Această poziție, susținută cu argumente de ordin filosofico-ontologic și etico-estetic, este prezentă și în secolul XX. Concepțiile despre om, care se remarcă printr-o reevaluare a corpului și a cărnii (vezi în special Nietzsche în filosofie, Gide în literatură ș.a.) folosesc datele oferite de dezvoltarea neurobiologiei, neuro-fiziologiei, psihanalizei, psihologiei, care evidențiază valoarea acestora pentru dezvoltarea echilibrată, normală a omului. Ele sunt completate cu teorii care încearcă să demonstreze rolul funcțional activ al gândirii și rațiunii în procesele biologice și psihologice. Poate fi citată în acest sens, teoria interacționistă schițată de K.R. Popper și J. K. Eccles<sup>4</sup>. În etică, A. Edel recomandă includerea persoanei în judecata etică, în unitatea structurilor sale, odată ce raționalitatea cuprinde toate nivelele individu-ality umane<sup>5</sup> în sprijinul acestei idei vin și teoriile moderne privind natura și rolul emoției în dezvoltarea personalității, așa cum se manifestă în cunoașterea și evaluarea lumii, în motivația, alegerea și manifestarea comportamentului. (Vezi teoria fiziologică, teoria cognitivistă, teoria conativă, teoria contextualistă, behaviorismul)<sup>6</sup>.

Aspectele menționate până în prezent privind relația dintre momentele cognitive și cele apreciative în formarea „imaginii eu“, se referă la importanța criteriilor de valorizare – oferite, în special, de punctul de vedere filosofic general asupra omului – în determinarea adevărului sau falsului cunoașterii de sine. Este vorba mai mult de o *atitudine culturală* a individului față de sine, cu efecte asupra gradului de cuprindere a sferei cunoașterii de sine și a ponderii factorului senzorial, afectiv și a celui rațional în autocunoaștere.

Un aspect al relației menționate, îl reprezintă întrepătrunderea dintre senzorial și rațional, dintre gândire și afectivitate în procesul formării „imaginii eu”. Relația este prezentă atât în structurile interioare ale eului, pe care „imaginea eu” le cuprinde ca momente integrate ale procesului constituirii sale, cât și în elementele separate ale structurii acesteia. Ponderea senzorialului și a raționalului este diferită în funcție de etapele procesului, ca și de formele sub care ele se manifestă.

Un prim moment al întrepătrunderii dintre senzorial și rațional îl reprezintă procesul cunoașterii de sine ca o descoperire a valorilor vitale. Prin valori vitale se înțelege, în acest context, sistemul exigențelor care exprimă și asigură unitatea, plenitudinea și dinamismul vieții, al unei trăiri intense, în care sunt dezvoltate potențele, capacitățile specifice individului și stimulată dezvoltarea și funcționarea armonioasă, normală a impulsurilor vitale – instincte și trebuințe biologice și psihologice fundamentale. Într-un cuvânt, valorile vitale exprimă bunăstarea și vigoarea organismului individual, sănătatea lui corporală, ca o condiție a celei spirituale. Prin însăși natura lor, ele sunt strâns legate de valorile morale. Referindu-se la raporturile de fundare reciproc indisolubile a valorilor vitale și a celor morale, Vasile Morar consideră că ele sunt impuse „în numele necesarei *solidarități active* a tuturor elementelor de ordin existențial cu omul, ca valoare primară, cu afirmarea vieții ca voință activă, sănătoasă și trăită în pace”<sup>7</sup>.

Cunoașterea de sine ca descoperire a valorilor vitale se exprimă, la un prim nivel – care nu este totdeauna depășit într-un nivel superior –, într-o conștiință vagă, imprecisă, neexprimabilă în limbaj, care privește afirmarea originară a unei spontaneități biologice constitutive, fiind mai direct legată de existența biologică a omului<sup>8</sup>. La un nivel superior, relația dintre valorile vitale și gândire se constituie pe planul unei conștiințe precise, exprimabile în limbaj. La acest nivel, relația menționată se realizează, pe de o parte, prin prezența conștientă a valorilor vitale ca *motivație* în modul de gândire, în raționalitate, ca modalitate de a semnifica și cuprinde în câmpul acțiunii raționale anumite domenii ale realității și ale raporturilor cu ceilalți, și, pe de altă parte, prin prezența lor ca *element* care oferă baza materială biofizico-psihologică a gândirii și, în același timp, ca factor constitutiv și tonus al acesteia.

În ceea ce privește raționalitatea sensibilității, ea constă la acest nivel, tocmai într-o anumită structură a valorilor vitale, construită prin intervenția gândirii, ca și în stimularea, pe baza unei cunoașteri raționale, a dezvoltării lor plene, a emancipării lor de orice împiedică manifestarea lor optimă. Atingerea acestui nivel al întrepătrunderii gândirii și a valorilor vitale, asigură calitatea superioară a fiecărui termen în parte, aflat în relație organică cu celălalt, ca și calitatea umană a individului, fiind un indicator al progresului realizat în umanizarea omului.

Un alt nivel important, elementar pentru formarea imaginii despre sine, îl constituie *schema corporală globală*, ca imagine a unității dintre părțile corpului și a relației lor în spațiu. Ea se află la baza activității motrice și constituie o condiție a oricărei experiențe. Mai mult decât atât, schema corporală determină o anumită viziune spațială a lumii, o imagine a ei determinată și o anumită relație spațială cu lucrurile exterioare, controlând poziția și experiența corpului propriu, contactul echilibrat cu realul. Ea nu reprezintă însă numai o condiție a echilibrului motric și senzorial, ci are o importanță mult mai însemnată, asigurând realizarea exigențelor instinctelor fundamentale, ale valorilor vitale, în raport cu condițiile existenței spațio-temporale. Așa cum precizează G. Gusdorf, ea exprimă sensul corpului propriu, posesia sinelui, care asigură securitatea intimă și sentimentul ei, reprezentând conștiința biologică globală<sup>9</sup>.

Pe lângă schema corporală definită mai sus, un alt element integrat al „imaginii eu” îl constituie *imaginea corporală estetică*, sau felul cum se vede individul în funcție de alcătuirea normală, sănătoasă sau nu a părților corpului său, a înfățișării sale estetice. Orice dezechilibru și dizarmonie a imaginii corporale în cele două sensuri menționate, stare patologică și anormală, infirmitate etc., produc efecte negative pe plan psihologic și moral: sentimentul frustrării, al neîncrederii în forțele proprii, timiditate – sau expresia ei camuflată manifestată în aroganță – izolare, anxietate și complex de inferioritate, cu toate consecințele ce decurg pentru comportament și relațiile interumane. Toate acestea constituie motive conștiente sau inconștiente ale conduitei, care poate prezenta un pronunțat caracter deviant, care poate ajunge de multe ori până la delincvență. Pentru a preîntâmpina fenomenele

negative, există încercări de a elabora mijloace de educare a conștiinței privind posibilitatea unei vieți demne chiar și în condițiile infirmității, ale unor defecte corporale care scad capacitatea fizică sau aspectul estetic, dar nu exclud compensarea prin educarea și manifestarea unor aptitudini spirituale deosebite<sup>10</sup>.

**Cunoaștere morală.** Problema întrepătrunderii elementelor senzoriale cu cele raționale, a celor cognitive cu cele apreciative, se complică la nivelul structurii „imaginii eu“, datorită complexității acesteia. Însăși „ imaginea eu“ reprezintă o unitate structurală dinamică de imagini, fiecare dintre ele având componentele sale cognitive și valorice. Dintre încercările de a le delimita merită oferită, ca exemplu, clasificarea lui M. Rosenberg, după care există „*eul real*“ (așa cum se vede individul în realitate, la un moment dat), „*eul dinamic*“ (așa cum și-a propus individul să devină), „*eul fantastic*“ (așa cum se vede individul în perspectiva însușirii normelor morale și a modelelor), „*eul viitor*“ sau „*eul posibil*“ (așa cum se vede individul ca tip de personalitate ce poate fi realizată), „*eul idealizat*“ (așa cum se vede individul pe sine cu plăcere), o serie de „imagini eu“ cu funcție de măști pe care individul le expune pentru a ascunde slăbiciunile eului real<sup>11</sup>.

În literatura de specialitate „ imaginea eu“ este analizată ca o formație socială, cu componentele sale cognitive, emoționale și comportamentale, care se constituie într-un sistem ierarhic de dispoziții: nivelul 1, format din *dispoziții elementare*, inconștiente, care se formează pe baza trebuințelor vitale, în situațiile cele mai simple și se manifestă în dispoziția spre acțiuni consolidate de experiența anterioară; nivelul 2 reprezentat de dispozițiile sociale, care se formează pe baza aprecierii situațiilor sociale; nivelul 3, constituit din *dispoziții sociale de bază*, care determină direcția generală a activismului individului; nivelul 4, superior, format din *sistemul orientărilor valorice*, spre scopurile și mijloacele activității, determinate de condițiile sociale generale de viață ale individului<sup>12</sup>.

S-a încercat și o delimitare a stadiilor formulării „imaginii eu“, precizându-se stadiile: 1. al *complexității și diferențierii cognitive*; 2. al *semnificației subiective pentru individ* a „imaginii eu“; 3. al *integrității, consecvenței interioare* a „imaginii eu“; 4. al *stabilității, fermității* „imaginii eu“<sup>13</sup>.

Din rezultatele cercetărilor de până acum reies unele particularități proprii „imaginii eu“. În primul rând, se remarcă atât complexitatea structurii sale și a factorilor care o determină, cât și caracterul său dinamic, deschis. În al doilea rând, se impune caracterul său contradictoriu. Posibilitatea și existența contradicțiilor este dată de complexitatea mecanismelor interacțiunii dintre diferitele nivele și elemente ale structurii „imaginii eu“, ale treceri unuia în celălalt, condițiile în care fiecare dintre ele își are autonomia sa funcțională și propriile sale contradicții, fiind alcătuite, la rândul lor, din nivele și elemente dinamic structurate, aflate în interacțiune și conflict posibil. Contradicțiile pot avea atât un rol disfuncțional, cât și unul de forță motrică. Astfel, anumite contradicții dintre componentele cognitive și cele emoționale ale eului, sau dintre interior și exterior, pot duce la disfuncții și forme patologice de manifestare a individualității. La același rezultat duc contradicțiile determinate de reprimarea trebuințelor vitale la nivelul cunoașterii de sine și a cunoașterii de sine morale, în virtutea unor concepții și modele care le valorizează negativ. În rândul contradicțiilor cu rol de forță motrică poate fi inclusă contradicția dintre eul real și cel ideal, tensiunea dintre ele determinând perfecționarea „imaginii eu“ în ansamblul său și creșterea efortului moral și a calității comportamentului corespunzător.

O particularitate a formării „imaginii eu“ ca proces specific de cunoaștere, o reprezintă prezența și importanța factorului subiectiv-afectiv-sentimente, emoții, autoaprecieri încărcate de subiectivism, dominate de dorințe, trebuințe, ambiții, impulsuri, interese personale. În aceste condiții, un loc important îl deține opinia, credința, certitudinea sensibilă. Autoaprecierea nu este deci bazată numai pe o *cunoaștere rațională*, adevărată sau falsă, ci și pe acțiunea unor instincte fundamentale, ca instinctul sau trebuința de afirmare și personalizare<sup>14</sup>. Acest lucru este firesc, în măsura în care autoaprecierea nu este numai cunoaștere și valorizare, ci și exigență, nu numai stabilirea unei coincidențe între cunoaștere și existență, ci și schițarea liniilor de depășire a existentului în posibil. Prin aceasta nu se neagă importanța elementului cognitiv în autoapreciere. Dimpotrivă, așa cum precizează I.S.Kon, autoaprecierile sunt scheme cognitive proprii, care

generalizează experiența trecută și organizează, structurează, noua informație cu privire la un anumit aspect al eului<sup>15</sup>. Dar, trebuie să adăugăm informația, organizarea ei, depinde de atitudinea emoțională față de sine, de respectul de sine, de tendința de a se vedea în lumina cea mai bună etc.

În concluzie, se poate afirma că îmbinarea, întrepătrunderea dintre elementul cognitiv și cel apreciativ, preponderența în unele stadii ale formării „imaginii eu“ a elementelor subiectivității, determină specificul cunoașterii ca fiind *cunoaștere morală*.

**Cunoaștere de sine și autoapreciere.** Pe ce căi se ajunge la descoperirea de sine și la autoapreciere? Dintre cele mai importante, care implică, în continuare, conștiința de sine morală, pot fi menționate câteva principale.

1. Există o experiență a trăirilor, a manifestării sensibilității, care își poate găsi momente de înaltă intensitate. O întâmplare dramatică în copilărie, o decepție în dragoste etc., pot marca puncte nodale în contactul cu sensibilitatea profundă și valorile care schițează atitudinea globală față de sine și față de viață în general. De aceea, în constituirea „imaginii eu“ un rol de importanță hotărâtoare îl deține *biografia individului*, cu evenimentele și experiențele sale decisive pentru orientarea în viitor, în atitudinea față de rostul și sensul vieții, față de momentele ei cruciale, față de moarte. Ele influențează, în grade diferite, conștient sau inconștient, comportamentul, exigența față de sine, în situațiile concrete.

2. Individul se descoperă, se cunoaște pe sine, în mod hotărâtor, în situațiile de viață și experiența cotidiană. El se confruntă pe sine cu cerințele care i se impun în activitatea zilnică, răspunde unor solicitări, este pus în situația de a se conforma unor norme, de a rezolva probleme. În aceste condiții, el își descoperă calități și defecte, compară succesele cu eșecurile, își verifică capacitățile și posibilitățile. Experiența cotidiană deține un rol important în modul de a fi al eului și al „imaginii eu“ în special, ca *unitate structurală complexă dinamică*. Relația dintre eul real și eul ideal sau eul viitor nu se constituie numai în funcție de ideal și aspirații, ci, în primul rând, ca un rezultat al verificării măsurii proprii, al capacității de a răspunde exigențelor vieții cotidiene. De asemenea, gradul în care se face apel la un „eu mască“, ca și numărul măștilor, depinde, cu deosebire, de relațiile cotidiene, de situațiile de viață concrete, în care individul este pus

în situația de a se conforma, manifestându-se în concordanță sau în contradicție cu eul său adevărat. G. Gusdorf surprinde acest fenomen atunci când stabilește diferența dintre eul profund și eul cotidian <sup>16</sup>. Într-adevăr, nu întotdeauna individul se poate manifesta deplin în experiența cotidiană, în concordanță cu eul său profund. Decalajul dintre ele poate fi determinat, printre altele, de neconcordanța dintre cerințele profunde, valorile vitale personale ale structurii intime proprii individului și cerințele impuse de familie, grup etc., care pot avea uneori un rol de reprimare. În asemenea situații, contradicția dintre eul profund și eul cotidian poate duce la o falsă cunoaștere de sine și o autocunoaștere denaturată ca forme ale conștiinței false sau la ipocrizie și duplicitate.

3. O altă cale prin care individul se cunoaște pe sine și se autoapreciază, își formează o imagine despre sine, constă în *relația sa cu altul*. În afara considerațiilor general filosofice și etice privind importanța și rolul *altuia* în autodezvoltarea individului, psihologia și psihologia socială oferă date concludente în explicarea științifică a relației eu-altul, eu-ceilalți, eu-noi. Ele trebuie încadrate în contextul cunoașterii relațiilor sociale și al rolului acestora în dezvoltarea existenței naturale a omului. Cu alte cuvinte, relația eu-altul, ca relație socială interindividuală, se integrează în ansamblul relațiilor sociale prin intermediul relațiilor umane.

În cadrul unor diferențe sau conflicte între interese, manifestate pe plan social, diferența dintre eul autentic și eul cotidian, așa cum se exprimă față de altul, se adâncește. Se creează, adesea, în aceste condiții, necesitatea și deprinderea afișării unor măști, care ascund eul autentic, influențând, prin purtare îndelungată, sau prin abilitatea formată de a le schimba, însăși autenticitatea acestuia, natura lui adevărată, ca și posibilitatea autocunoașterii și a cunoașterii lui de către altul și de ceilalți. În procesul conștient sau inconștient de elaborare a unor „euri-măști“ utile, eficiente în realizarea unor scopuri impuse de o situație de viață concurențială, se constituie conștiința falsă despre sine și o „imagine eu“ falsă. Distanța și înstrăinarea dintre eu și altul, duc astfel, la îndepărtarea și înstrăinarea eului de sine însuși, ca modalitate și formă esențială a înstrăinării de sine. Eul-mască, se substituie, astfel, eului real, exprimând un aspect fundamental al conștiinței false, în condițiile unor raporturi sociale și *interumane* bazate pe interese antagoniste.

Dimpotrivă, în cadrul unor relații bazate pe interese mutuale, care fac posibile și necesare raporturi echitabile între indivizi, reglementate de principiul echității, relația transparentă, sinceră, bazată pe respect reciproc și încredere, exigență și atitudine constructiv-critică față de altul devine o *necesitate morală*. În măsura în care înflorirea tuturor capacităților creatoare ale individului reprezintă o condiție esențială pentru participarea sa optimă la realizarea unor scopuri și interese *comune*, auto-cunoașterea și cunoașterea *adevărată a celuilalt* constituie o *necesitate socială*. În aceste condiții, este posibilă și necesară dezvoltarea eului, concordanța dintre nivelele sale, ca și dintre diferitele elemente ale structurii „imaginii eu“.

Pe fondul social general, se înscriu modalitățile specific psihologice ale relației dintre eu și altul, care contribuie la formarea „imaginii eu“. Ele au fost studiate și delimitate atât de psihologie, cât și de psihologia socială și sociologie. De o importanță deosebită s-a dovedit a fi părerea altuia în procesul auto-aprecierii, ca aspect al formării „imaginii eu“. Ea influențează atât reprezentarea modului în care eul apare în ochii altuia, este apreciat de altul, cât și autoaprecierea, însoțită de sentimentul mândriei, rușinii, umilinței etc., în funcție de raporturile cu altul și părerea altuia.

Orientarea interacționistă inițiată de G. H. Mead în psihologia socială, stabilește relația dintre *eu* și *noi*, constituită într-o practică comună, în care „noi“ reprezintă altul generalizat. Auto-reflecția reprezintă, în această lumină, capacitatea de a se situa în locul altuia fapt ce presupune nu numai înțelegerea și cunoașterea altuia, dar și identificarea cu altul.

Relația eu-altul este importantă pentru modul și gradul cunoașterii de sine, cât și în procesul și calitatea autoaprecierii. Dorința de a fi bine văzut și apreciat de altul, ca o condiție a prestigiului personal, a afirmării valorii proprii în fața altuia, constituie un imbold pentru dezvoltarea acelor însușiri care au mai mult preț în ochii acestuia. Exigența față de sine, ca și încrederea în sine și chiar alegerea de sine, sunt, deci, în mare parte, determinate de părerea altuia. Într-un fel, altul are rol de mediator între mine și mine, atât în ceea ce privește procesul cunoașterii și aprecierii de sine – mă văd cu ochii altuia – cât și ca arbitru în judecata între ceea ce sunt și ceea ce pot sau trebuie să fiu.



4. Psihologia socială descoperă o altă cale prin care individul își creează „imaginea eu” și anume interiorizarea rolului social. Descrierea dialecticii individualului și socialului prin intermediul *rolului* se realizează pe trei nivele: sistemul impersonal macro-social (nivelul sociologic); interacțiunea interpersonală directă (nivelul socio-psihologic); motivația individuală. Toate nivelele, în interacțiunea lor, fiecare cu o pondere mai mare sau mai mică, influențează gradul și calitatea *interiorizării rolului* ca parte a „imaginii eu”. Individul se apreciază, își acordă o valoare determinată și, sau chiar în primul rând, după rolul pe care îl deține în societate, prestigiul social al acestuia etc. Rolul are, în acest caz, funcția de criteriu al autoaprecierii în ansamblul său. El este, în același timp, un criteriu de selectare a însușirilor necesare pentru îndeplinirea sa optimă, cu alte cuvinte, reprezintă un factor de autodezvoltare a „imaginii eu”. Forța dinamizatoare a interiorizării rolului crește în cazul în care există un conflict între diferite roluri ale individului, sau alternative de opțiuni între roluri diferite. În această situație, individul este obligat să-și cunoască bine capacitățile, să-și compare aptitudinile, să se aleagă pe sine, alegând între roluri. Intervine în acest caz responsabilitatea pentru afirmarea valorii proprii și alegerea mijloacelor corespunzătoare. Rolul poate determina însă și stări conflictuale cu efect negativ, care pot duce la contradicții între diferitele imagini constitutive ale structurii „imaginii eu”, la înstrăinare de sine, duplicitate etc. Aceste disfuncții și contradicții ale „imaginii eu”, cu consecințe în comportament, sunt prezente, cu deosebire, în cazul unui rol nepotrivit, nedorit, străin de aptitudinile, aspirațiile și idealul individului.

În structura subiectivității, „*imaginea eu*” reprezintă *nivelul de trecere de la conștiința de sine, la conștiința de sine morală*, deoarece îmbină, în procesul formării și manifestării sale, momentele cognitive cu cele valorizatoare, ultimele presupunând o apreciere cu conținut moral. Ea nu mai este numai o cunoaștere de sine, o simplă constatare a unui fapt dat, o dată ce conține în sine stimulentele și forțele necesare pentru o dezvoltare pe linia satisfacerii unor exigențe psihologice sau morale, a unui mod dorit de a fi sau a apărea în fața celorlalți. Ea nu este încă o conștiință de sine morală, pentru că elementele valorizatoare privesc *valoarea individului* – în ultimă instanță și în general – *prin prisma interesului său personal exclusiv, autoaprecierea fiind o valorizare pentru sine*.

Trecerea de la „imaginea eu“ la conștiința de sine morală se realizează prin ridicarea exigențelor formării unei „imagini eu“ adecvate, la situația morală elementară. Reiese, că trecerea menționată, *poate să nu aibă loc*, autoaprecierea și valorizarea să se oprească la nivelul „imaginii eu“. *Conștiința de sine morală este rezultatul unei conștiințe morale proprii persona-lității, care presupune problematizarea și controlul auto-cunoașterii în lumina situației morale elementare.* Prin situație morală elementară, proprie unei societăți democratice, înțelegem asumarea de angajamente și responsabilități pentru valorizarea și manifestarea plenară a omului ca scop ultim și valoare fundamentală a societății.

## NOTE

<sup>1</sup> I. S. Kon, *Descoperirea eului*, Moscova, 1978, p. 9.

<sup>2</sup> *Ibidem*, p. 10.

<sup>3</sup> Vezi Maurice Merleau-Ponty, *L'homme et l'adversité*, în *La connaissance de l'homme au XX-ème siècle*, Editions de la Braconniere, Neuchâtel, 1952, p. 52.

<sup>4</sup> Vezi R.K. Popper, J.K. Eccles, *The Self And Its Brain* (1977), în care subliniază influența inversă a conștiinței asupra proceselor neuronale.

<sup>5</sup> Vezi A. Edel, *Ethical Judgement. The Use Of Science In Ethics*, The Free Press of Glencoe, New York, 1964, p. 55.

<sup>6</sup> Cf. Ronald de Sousa, *The Rationality Of Emotion*, The MIT Press, Cambridge, Massachusetts, Londra, 1990, pp. 36-46.

<sup>7</sup> Vasile N. Morar, *Etica și mutațiile valorice*, Editura Științifică și Enciclopedică, București, 1985, pp. 66-67.

<sup>8</sup> Vezi, în această problemă, Georges Gusdorf, *La découverte de soi*, P.U.F., Paris, 1948, p. 301 și urm.

<sup>9</sup> Vezi G. Gusdorf, op. cit., p. 353, 2. Foarte interesante pentru formarea „imaginii eu“ sunt cazurile patologice cinestezice, când locul normal al unor părți corporale este senzorial schimbat: fruntea se simte în locul ochilor, fesele în locul capului etc. Aceste stări patologice produc mari tulburări în relația și echilibrul cu lumea exterioară. La fel de relevante pentru cunoașterea caracterului bio-fizico-psihologic al schemei corporale, sunt cazurile în care imaginea eului corporal în integritatea sa persistă, chiar atunci când una din părțile sale lipsește, de exemplu perceperea unui membru amputat (membrul-fantomă), sau „amputarea psihologică“, când nu există conștiința și nu este perceput un element corporal existent (mână, ochi etc.).

<sup>10</sup> Există în țările anglo-saxone, în special în America, o medicină morală – Christian Science – care încearcă să educe în spiritul acceptării și depășirii stărilor patologice sau anormale, determinate de infirmități corporale. Despre aceste probleme, vezi P. Janet, *Médications Psychologiques*, vol. 1; Dr. Paul Tournier, *Médecine de la Personne*, Delacheux et Niesle, 1941, citați de G. Gusdorf, *op. cit.* p. 359.

<sup>11</sup> M. Rosenberg, *Society and the Adolescent Self-Image*, Princeton, 1965, citat de I. Kon, *op. cit.* pp. 65–66.

<sup>12</sup> Vezi I. S. Kon, *op. cit.*, pp. 68–69.

<sup>13</sup> *Idem*, pp. 70–71.

<sup>14</sup> Dintre formele patologice de manifestare a personalității, pot fi citate *depersonalizarea*, care dereglează aspectele emoționale ale conștiinței de sine, dar nu tulbură claritatea celor cognitive, manifestându-se în sindromul automatismului psihic, în care senzațiile, mișcările etc., par străine; *dedublarea personalității* sau *pluralitatea eului*, când apare dublura.

<sup>15</sup> I. S. Kon, *op. cit.*, p. 55.

<sup>16</sup> Vezi G. Gusdorf, *op. cit.*, p. 393.

# PROBLEMA UNITĂȚII ȘTIINȚEI ȘI INTELIGENȚA ARTIFICIALĂ

Mircea MALIȚA și Mihaela MALIȚA

(Academia Română, București)

1. Filosofia luminilor ne-a lăsat moștenire ideea *unității științei*. Unitatea este prima din trăsăturile ce caracterizează de obicei știința: unitate, generalitate, certitudine, obiectivitate, fundamentare metodică, dezvoltare progresivă, predicție. Știința a fost considerată unitară în secolul luminilor pentru că o singură rațiune și un singur demers rațional permiteau cunoașterea fenomenelor naturii. Baza unității științei era astfel *epistemologică*. Omul și societatea făcând parte din natură, deosebiri între fenomene nu păreau să tulbure acest demers, mai ales că dualismul cartezian și disfuncția între spirit și materie erau în declin. Deci unitatea științei avea și un fundament *ontologic*. Consecințele asupra formării și dezvoltării științelor sociale și umane au fost mari: ideea unității științei legitima modelul metodelor newtoniene pentru domeniile în curs de emancipare de sub tutela tradițională: științe juridice, științe istorice, științe sociale și psihologice. Toate au adoptat de la începuturile lor limbajul științei mecanice: forțe, cauze, legi, ipoteze, verificare experimentală.

Raționalistii și empiriștii au concurat în susținerea tezei unității științei. Descartes scrisese că „științele sunt în prezent mascate și e de ajuns să le ridicăm masca pentru ca înlănțuirea lor să fie tot așa de ușor de reținut ca și seria numerelor”. Iar școala empiristă ilustrată de David Hume și John Stuart Mill susținea că logica formării conceptelor, construirea teoriilor și verificarea ipotezelor sunt comune tuturor științelor. Diferențe pot exista în procedurile de a descoperi faptele relevante (contextul descoperirii), dar nu în metodele de explicație (contextul justificării). Științele naturale și sociale sunt astfel legate, în principal, după empiriști, de o unitate *metodologică*.

2. Este ușor de înțeles de ce în secolul al XIX-lea apare ontestarea tezei unității științei sub forma curentului care subliniază specificul fiecărei științe și diferențele ireductibile între ele. Științele sociale resimțeau constrângerile și limitele metodelor dominante în științele naturii și pledau pentru metode proprii.

Nu este suficient să observi din afară o structură socială ca pe un fapt fizic, spuneau sociologii, ci este necesar să vezi cum ea înțelege realitatea și să interpretezi modul în care se conceptualizează. De aceea, poziția lor a fost numită *interpretaționism*. Cercetarea socială a început să se delimiteze de științele naturii atât în ce privește explicația cât și metodele. Max Weber cu conceptul său „Verstehen” și Émile Durkheim cu efortul său de a construi o știință socială cu concepte teoretice proprii se înscriu în această tendință. Putem adăuga pe R. G. Collingwood cu a sa „metodă istorică”, conform căreia faptele istorice nu pot fi investigate cu metodele științelor naturale și pe Wilhelm Dilthey, care a stabilit o autonomie metodologică pentru „Geisteswissenschaften”.

Următorul pasaj al lui Dilthey, filosoful care a spus că „noi explicăm natura, dar înțelegem viața psihică”, este concludent pentru afirmarea de către el a breșei între științele naturii și științele omului, simultan cu susținerea unității celor din urmă: „Alături de științele naturii s-a dezvoltat spontan pornind de la sarcinile vieții însăși, un grup de cunoștințe care sunt legate de un obiect comun. Acestea sunt științe ca istoria, economia politică, științele juridice și politice, arhitectura și muzica, concepțiile asupra lumii și sistemele filosofice, în sfârșit, psihologia. Toate aceste științe se ocupă de unul și același fapt: genul uman. Ele descriu și relatează, formează judecăți, concepte și teorii referitoare la acest fapt... Astfel apare posibilitatea de a determina acest grup de științe prin relația lor comună cu unul și același fapt – umanitatea – și de a le separa de științele naturii.”

Cu toate curențele care subliniază diversitatea și specificitatea științelor, teza unității științei este reafirmată în secolul XIX de Karl Marx atât pe baze ontologice (unitatea materiei), cât și epistemologice (metoda dialectică).

Este de amintit că Auguste Comte, deși inițial a susținut că între științe sunt bariere de netrecut („vom privi ca temerar să

aspirăm vreodată, chiar pentru viitorul cel mai îndepărtat, a le reduce la una singură“ (vol. I.), în *Cursul de Filosofie pozitivă* (ultimul volum) urmărea să arate toate științele ca ramuri ale unui trunchi unic.

3. În secolul nostru, momentul cel mai patetic al proclamării unității științei l-a constituit școala neopozitivistă de la Viena. Congresul de la Paris, organizat în principal de protagoniștii acestei școli în 1935 s-a numit chiar Congresul Internațional pentru Unitatea Științei și și-a propus elaborarea unei enciclopedii internaționale a științei unitare. După Rudolf Carnap, „... analiza logică, cu procedurile logicii noi, duce la o *știință unitară*. Nu există diferite științe, cu metode esențial diferite sau cu izvoare cu totul diferite de cunoaștere; nu există decât Știința. Toate cunoștințele găsesc aici locul lor și aceste cunoștințe sunt toate de aceeași natură. Diversitatea lor aparentă nu este decât efectul diversității limbajelor întrebuintate în părțile științei“.

Unificarea științelor ar presupune adaptarea limbajului științelor fizice ca limbaj universal, ceea ce a dat numele de *fizicalism* acestui curent. Critica acerbă pe care au adus-o deceniile următoare neopozitivismului și reduccionismului (reducerea tuturor fenomenelor la cele fizice) explică, și îndoiala cu care a fost privit proiectul unificării, așa cum a fost conceput în 1935.

Această sumară incursiune istorică ne permite să înțelegem că diferitele științe în dezvoltarea lor au pendulat între nevoia de a sublinia fie caracterul unitar al tuturor științelor, fie caracterul distinct și autonom al fiecărei științe. În același timp aruncă o lumină asupra situației actuale.

4. Este îndeobște cunoscută poziția cardinală pe care la mijlocul secolului nostru au dobândit-o cibernetica, teoria sistemelor, modelarea matematică și abordarea interdisciplinară, precum și înrâurirea lor asupra dezvoltării tuturor științelor. Aceste teorii și școli, care ca și disciplinele matematicii, logicii, semioticii, praxeologiei, au vocație unificatoare, se prevalează într-o formă sau alta de ideea unității științei. Cibernetica a scos în relief procesele comune de comunicare și control care există în natură, societate și organisme vii, justificând astfel tratarea lor cu aceeași familie de metode. Folosirea modelelor matematice presupune legitimitatea transferului lor dintr-un domeniu de cercetare în altul, iar interdisciplinaritatea privește granițele dintre științe ca niște delimitări convenționale de arii, orice metodă născută într-o arie putând fi preluată și însușită în alta.

În ce privește teoria sistemelor, ea își propune direct și deschis, obiectivul de a oferi o unificare a științei pe noi baze, diferite de cele propuse de mecanismul secolelor al XVIII-lea și al XIX-lea sau de fizicismul mai recent. „Nu putem reduce toate nivelele biologice, comportamentale și sociale, la cel mai de jos nivel, acela al construcțiilor și legilor fizicii... Principiul unificator constă în faptul că găsim *organizarea* la toate nivelele“, scria L. von Bertalanffy. „Suntem cu siguranță capabili, să stabilim legi științifice pentru diferitele nivele sau straturi ale realității. Și aici găsim... o corespondență sau izomorfie a legilor și schemelor conceptuale în diferite științe, garantând unitatea științei“. Viziunea sistemelor asupra lumii se pronunță astfel explicit în favoarea unității: „Realitatea, în concepția modernă, apare ca o uriașă ordine ierarhică de entități organizate, care duc, într-o suprapunere de multe nivele, de la sistemele fizice și chimice la sistemele biologice și sociologice. Unitatea științei e garantată, nu de o utopică reducere a tuturor științelor la fizică și chimie, ci de uniformitățile structurale ale diferitelor nivele ale realității.“

„Lumea noastră e o lume de sisteme“, va preciza Bunge. „Numim *sistemică*, scrie el, mulțimea teoriilor care se ocupă de caracteristicile structurale ale sistemelor și care pot deci trece barierele mai degrabă artificiale dintre discipline“.

5. Cel mai recent capitol al culturii științifice dominante în vremea noastră după cibernetică, sisteme, modele, interdisciplinaritate este inteligența artificială. Din primele sale tentative de a modela operațiunile inteligente pe calculator se degajă afinitatea cu premisele celorlalte domenii amintite. „Se poate vedea că această abordare, scriu Newell și Simon, nu face nici o presupunere că structurarea computerelor și a creierelor ar fi similare, în afara presupunerilor că ambele sunt mașini cu scopuri generale de manipulat simboluri și că, un calculator poate fi programat să execute procese elementare de informație funcțional asemănătoare cu cele executate de creier“. În această afirmație, alături de izomorfismul modelelor, despre care am vorbit mai sus, este cuprinsă caracterizarea minții ca proces de simboluri cu scop general. Într-adevăr, prima fază a inteligenței artificiale caută procedee generale, aplicabile la raționamentele și problemele oricărui domeniu de investigație și deci ale întregii științe, privite ca un tot unitar. În 1957, programul numit LOGIC THEORIST (LT) al lui

Newell, Simon și Shaw a demonstrat 38 de teoreme din *Principia Mathematica*. (Celelalte teoreme au fost demonstrate ulterior tot pe cale automată de Wang, dar cu o altă metodă). *Rezolvatorul general de probleme* (GPS) al lui Newell, Simon și Shaw reflectă ca și predecesorul său LT adeziunea la o știință a cărei unitate rezidă în unicitatea metodei. Prima lor ipoteză de lucru este: „Un sistem de simboluri fizice posedă mijloacele necesare și suficiente pentru acțiune inteligentă generală“. A doua ipoteză este aceea a cercetării euristice. „Soluțiile la probleme sunt reprezentate ca structuri de simboluri. Un sistem de simboluri fizice exercită inteligență în rezolvări de probleme prin *căutări*, adică generând și modificând progresiv structuri de simboluri până când produce structura soluției“.

Optimismul manifestat în 1957 la nașterea GPS nu a fost confirmat de aplicațiile practice ale metodei generale și unice. Simon prezisese în acel an că în 10 ani: a) un computer ve deveni campionul mondial de șah; b) un computer va descoperi și demonstra o teoremă matematică nouă și c) cele mai multe teorii în psihologie vor avea forma programelor de calculator. Deși programele de inteligență artificială au progresat mult în domeniile amintite, suntem încă departe de performanțele prezise (cu toate că peste 25 ani mașinile de șah din comerț băteau 99% din amatorii de șah).

6. Este interesant de semnalat faptul că cele mai mari progrese ale inteligenței artificiale au fost realizate după părăsirea ipotezei unui singur procedeu unic, în favoarea unor procedee specifice fiecărui domeniu al cunoașterii și al științei. Aceasta este faza „sistemelor expert“.

Sistemele expert sunt programe care simulează capacitatea de a rezolva probleme a unui expert într-un domeniu (medicină, geologie, construcții etc.). Ele cuprind a) faptele domeniului b) regulile, dominant euristice, ale practicii și raționamentului, specifice acestui domeniu, așa cum sunt folosite de un expert uman. Sistemele expert consistă dintr-o bază de cunoștințe și o procedură de inferență. Întrucât *cunoștința* este baza acestor sisteme, problemele construirii lor sunt reprezentarea cunoașterii, utilizarea și achiziția ei. Un exemplu îl constituie MYCIN – un program de diagnostic medicală și terapie în bolile infecțioase, în special ale sângelui și meningitei. Programul are 500 de reguli constând din formula „dacă... atunci“, prima parte definind condiții care, întrunite, fac să urmeze o concluzie. În alte sisteme



expert, cunoașterea este reprezentată de rețele semantice. În toate sistemele expert, există o scară a credibilității faptelor regulilor: de exemplu 1 este certitudinea, 0,9 evidența puternică, 0,6 evidența sugestivă etc. Felul de acțiune al programului este o secvență de reguli care concludă din teste și din istoria pacientului, prezența unei maladii.

„Medicina este probabil cel mai mare câmp neutilizat de aplicații pentru tehnicile deciziei, atât tradiționale, cât și cele bazate pe cunoaștere“ a spus un specialist.

Sistemul expert MYCIN este în fond un sistem de decizie, dar spre deosebire de tehnicile tradiționale, de exemplu un model de decizie care pe baza simptomelor (evidențe sau fapte) oferă o concluzie (indicarea maladii), are următoarele caracteristici remarcabile: permite interacțiunea utilizatorului cu calculatorul și comunicarea în limbaj natural; sistemul oferă explicații asupra deciziilor sale; este versatil (face lucruri în moduri diferite); este flexibil (recunoaște o orientare greșită și recuperează); învață; folosește cunoaștere explicită (spre deosebire de modelele cantitative tradiționale).

Înflorirea sistemelor în medicină (sistemul INTERNIST controlează 500 de boli cu 35 000 simptome), în chimie (DENDRAL decide structura chimică pe baza datelor fizice – chimice), în detectarea zăcămintelor de petrol, în finanțe, management etc. este la ordinea zilei.

Caracterizarea tendinței actuale de a crea programe și sisteme inteligente se rezumă la faptul că ipoteza fundamentală a anilor '60, anume că există o singură schemă de funcționare a cunoașterii și gândirii, este părăsită în favoarea tezei, că fiecare problemă trebuie judecată în Contextul unei situații specifice și pe baza cunoștințelor și regulilor proprii unui domeniu, teză care subminează unitatea metodologică a științei.

7. Nu înseamnă că succesul acestei orientări a eliminat cu totul tendința metodologiei unice semnalată anterior. În 1984, după cinci ani de cercetări, Simon a revenit pe scenă, și împreună cu Bradshaw și Langley a lansat programul BACON, apt să degajeze legi matematice dintr-o mulțime de date și să detecteze deci regularitățile în procese. Astfel, fiind confruntat cu datele de care dispunea Kepler asupra distanțelor dintre planete și a mișcărilor lor, programul a produs legea a 3-a a lui Kepler:  $P = KD^{3/2}$  (unde P este perioada și D distanța de soare). Cu același aparat

euristic programul a stabilit, pe baza datelor experimentale corespunzătoare, legea lui Ohm privind curentul electric și legea lui Black referitoare la propagarea căldurii. Legile obținute cu unul și același program aparțin unor domenii depărtate ale științei. Simon crede că „BACON oferă o ipoteză asupra felului în care se face o descoperire științifică“. Procesele pe care BACON le folosește sunt cele care rezultă din alte cercetări asupra rezolvării problemelor și care sunt descrise ca procese de căutări euristice selective în spații de soluții posibile“. Pentru prezentul articol, nu atât dezbaterea aprinsă care a urmat asupra pretenției modelului BACON de a arunca o lumină asupra genezei teoriilor științifice este relevantă cât faptul că din nou euristica folosită este independentă de domenii, sugerând o metodologie comună și o știință unitară.

Reflecția pe care o propunem în concluzie este următoarea:

- istoria îndelungată a științei demonstrează existența unei oscilații între tezele extreme ale unității și diversității științei, care joacă rolul a doi poli;
- filosofia științei se apropie de un rol sau altul, după cum științele au nevoie pentru dezvoltarea lor de un teren comun, care să propage metodele de la științele care înregistrează succesele cele mai mari la științele mai puțin avansate sau simt necesitatea elaborării unei metodologii proprii;
- scurta istorie a inteligenței artificiale manifestă și ea o oscilație între ipoteza procedurii unice sau a procedurilor diverse ale operațiilor inteligente, iar pendularea este determinată în mare măsură de succesul practic al aplicațiilor permise de o abordare sau alta.

## BIBLIOGRAFIE

1. L. von Bertalanffy, *General System Theory* G. Braziller, New York, 1968.
2. Bradshaw G.F, Langley P.W., Simon H.A., *Studying Scientific Discovery by Computer Simulation*, Science 2, Dec. 1983.
3. M. Bunge, *Treatise on Basic Philosophy*, Vol. IV, A World of Systems, Reidel, Amsterdam, 1979.

4. R. Carnap, *L'ancienne et la nouvelle logique*. Hermann, Paris, 1933.
5. E. Cassirer, *Descartes et l'unité de la science*, Revue de Synthèse XIV, 1, April 1937.
6. *Computer Culture*: Annals of N.Y. Academy of Sciences, Ed. H. Pegels, 1984.
7. A. Comte, *Cours de Philosophie positive*. Paris 1830–1842, Ed. 2, 1864.
8. W. Dilthey, *Gesammelte Schriften* (vol. VII. Grundlegung der Geisteswissenschaften) Goettingen, Vandenhoeck & Ruprecht, 1913–1967.
9. H. Dreyfus, *What Computers Can't Do*. The Limits of A.I. Harper, 1972.
- 10 J. Fox, *Expert systems and the concept of Knowledge in Robots and A.I.* Eds. Brady, Gerhardt, Davidson, Springer, Berlin, 1984.
11. J.V. Grabiner, *Computers and the nature of man: a historian's perspective on controversies about A.I.* Bull. Am. Math. Soc. Oct. 1986.
12. J. Habermas, *Connaissance et intérêt*. Gallimard, 1976.
13. A. Levison, *Knowledge and Society*, Pegasus Indianapolis, 1974.
14. A. Newell, H. Simon, *Human Problem Solving*, Prentice – Hall, Englewood Cliffs, N.Y., 1972.
15. A. Posescu, *Logica Științei*, București, 1942.



# PARTEA A II-A PROBLEMA *MIND-BODY* ȘI EXPERIMENTUL MENTAL

## PROBLEMA MINTE-TRUP \*

Karl R. POPPER

(Londra)

### I. LUMEA A 3-A ȘI PROBLEMA MINTE-TRUP

Una din ipotezele centrale în *The Self and Its Brain* este aceea că problema minte-trup poate apărea într-o lumină nouă din perspectiva lumii a 3-a. Am să enunț pe scurt trei argumente în acest sens.

Primul argument este următorul:

(1) Obiectele lumii a 3-a sunt abstracte (chiar mai abstracte decât forțele fizice) dar nu mai puțin reale; ele sunt unelte puternice pentru transformarea lumii. (Nu vreau să sugerez că acesta este singurul motiv pentru a le considera reale sau că ele nu sunt altceva decât unelte).

(2) Obiectele lumii a 3-a au un efect asupra lumii 1 numai prin intervenția oamenilor, intervenția celor care le-au creat mai ales prin faptul de a le fi înțeles (*grasped*), care este un proces al lumii a 2-a, un proces mental, sau mai precis un proces în care interacționează lumea a 2-a cu lumea a 3-a.

(3) Prin urmare trebuie să admitem că atât obiectele lumii a 3-a cât și procesele lumii a 2-a sunt reale – chiar dacă s-ar putea să nu ne placă acest lucru din respect, să zicem, față de marea tradiție a materialismului.

Cred că primul argument este acceptabil – deși se poate nega, desigur, oricare din asumțiile sale. Se poate nega, de pildă, faptul că teoriile sunt abstracte, sau că au ca efect asupra lumii 1, sau chiar pretenția că teoriile abstracte pot afecta direct lumea fizică. (Cred totuși că oricare din aceste critici ar fi dificil de susținut).

Al doilea argument depinde parțial de primul. Dacă admitem interacțiunea dintre cele trei lumi și astfel realitatea lor, atunci

---

\* *The Mind-Body Problem, Popper's Selection*, David Miller (ed.), Princeton, Princeton University Press, 1985.

interacțiunea dintre lumea a 2-a și lumea a 3-a, pe care, într-o anumită măsură, o putem înțelege, ne poate ajuta să înțelegem mai bine și interacțiunea dintre lumea 1 și lumea a 2-a, o chestiune care face parte din problema minte-trup. Pentru că una din interacțiunile existente între lumea a 2-a și lumea a 3-a (*înțelegerea*) poate fi interpretată ca o fabricare de obiecte ale lumii a 3-a și ca o combinare a lor prin selecție critică; același lucru pare să fie valabil și în cazul percepției vizuale a unui obiect al lumii 1. Aceasta sugerează că ar trebui să considerăm lumea a 2-a ca fiind activă – productivă și critică (fabricare și combinare). Dar avem motive să credem că același lucru este realizat și de unele procese neurofiziologice inconștiente. Astfel, devine mai ușor de *înțeles* faptul că procesele conștiente pot acționa de-a lungul unor coordonate similare: este '*de înțeles*', până la un punct, faptul că procesele conștiente realizează sarcini similare celor realizate de procesele nervoase.

Al treilea argument în legătură cu problema minte-trup se referă la statutul limbajului uman.

Capacitatea de a învăța un limbaj – și chiar nevoia puternică de a învăța un limbaj – aparțin, se pare, structurii genetice a omului. Pe de altă parte, învățarea efectivă a unui limbaj particular, deși este influențată de nevoi și motive innăscute inconștiente, nu este un proces reglat genetic și deci nu este un proces natural, ci un proces cultural, un proces al lumii a 3-a. Astfel, învățarea limbajului este un proces în care dispozițiile genetice, dezvoltate prin selecție naturală, se suprapun și interacționează cu un proces conștient de învățare și explorare, bazat pe evoluție culturală. Acest fapt susține ideea unei interacțiuni între lumea 1 și lumea a 3-a și, din perspectiva argumentelor anterioare, susține existența lumii a 2-a.

Relația dintre evoluția genetică și evoluția culturală a fost discutată de o seamă de biologi eminenți<sup>1</sup>. Putem spune că evoluția culturală continuă evoluția genetică cu alte mijloace: cu ajutorul obiectelor lumii a 3-a.

Se spune adesea, în mod justificat, că omul este un animal care fabrică unelte. Dacă prin unelte înțelegem corpuri materiale, atunci este foarte important să observăm că nici una dintre uneltele umane nu este determinată genetic, nici măcar bățul.

Singura unealtă care pare să aibă o bază genetică este limbajul. Limbajul este non-material și apare în cele mai variate forme fizice – adică în forma unor sisteme foarte diferite de sunete fizice.

Unii behavioriști preferă să nu vorbească de „limbaj” ci numai de vorbitori ai unui limbaj particular. Și totuși limbajul înseamnă mai mult decât atât. Toți oamenii normali vorbesc; și acest lucru este extrem de important pentru ei; atât de important încât până și o fetiță surdă, mută și oarbă ca Helen Keller a dobândit, cu entuziasm și rapiditate, un substitut pentru vorbire prin intermediul căruia a ajuns să stăpânească efectiv limba engleză. Din punct de vedere fizic, limbajul ei era foarte diferit de engleza vorbită; dar avea o corespondență de unu la unu cu engleza scrisă sau tipărită. Fără îndoială că ea ar fi putut dobândi în același fel oricare altă limbă în afara englezei. Ea resimțea, deși în mod inconștient, nevoia de limbaj – limbaj în sens abstract.

Așa cum reiese din numărul și varietatea lor, limbajele sunt fabricate de om: ele sunt obiecte culturale ale lumii a 3-a, deși posibilitatea lor depinde de abilități, nevoi și țeluri care au devenit o moștenire genetică. Fiecare copil normal dobândește un limbaj printr-o muncă activă intensă, plăcută, poate și dureroasă. Achiziția intelectuală obținută astfel este însă imensă. Acest efort are un puternic efect feedback asupra personalității copilului, asupra relațiilor sale cu alte persoane și cu mediul său material.

Astfel, putem spune că într-o anumită măsură copilul este produsul acestei realizări proprii. El însuși este, parțial, un produs al lumii a 3-a. Așa cum adaptarea copilului la mediul său material și conștiința sa despre această adaptare sunt extinse prin abilitatea nou dobândită de a vorbi, același lucru se petrece și cu conștiința sa de sine. Eu-l, personalitatea copilului, apare în interacțiune cu alte eu-ri, cu artefactele și celelalte obiecte din mediul său. Toate acestea sunt profund afectate de însușirea limbajului; mai ales atunci când copilul devine conștient de numele său, când învață să denumească diferitele părți ale trupului și, cel mai important, când învață să folosească pronumele personal.

A deveni o ființă umană normală depinde de un proces de maturizare în care dobândirea limbajului joacă un rol extrem de important. Copilul învață nu numai să perceapă și să-și inter-

preteze percepțiile, dar învață și să fie o persoană, să fie un eu. În acest sens, concepția potrivit căreia percepțiile noastre ne sunt „date“ este, cred, greșită: ele sunt „fabricate“ de noi, sunt rezultatul unei munci active. La fel de greșită consider și ignorarea faptului că celebrul argument cartezian „gândesc, deci sunt“ presupune limbajul și abilitatea de a folosi pronumele (ca să nu mai vorbim de formularea problemei, foarte sofisticată de altfel, pe care se presupune că argumentul o rezolvă). Kant pare să nu se fi gândit la starea pre-lingvistică sau pre-filosofică a unui copil (sau a lui însuși) atunci când spune că acest „eu gândesc“ însoțește toate percepțiile și experiențele noastre <sup>2</sup>.

## II. MATERIALISMUL ȘI AUTONOMIA LUMII A 3-A

Cum arată însă lumea a 3-a dintr-un punct de vedere materialist? Existența avioanelor, aeroporturilor, bicicletelor, cărților, clădirilor, mașinilor, computerelor, gramofonelor, cursurilor, manuscriselor, picturilor, sculpturilor și telefoanelor nu prezintă o dificultate pentru nici o formă de fizicalism sau materialism. În timp ce pentru un pluralist acestea sunt exemplificări materiale, întrupări ale unor obiecte ale lumii a 3-a, pentru materialist ele sunt pur și simplu părți ale lumii 1.

Dar ce spune materialismul în legătură cu relațiile logice obiective care există între teorii (scrise sau nu), cum ar fi incompatibilitatea, deductibilitatea reciprocă, suprapunerea parțială, etc.? Materialistul radical înlocuiește obiectele lumii a 2-a (experiențe subiective) cu procese ale creierului. Între acestea, de o deosebită importanță sunt dispozițiile pentru comportamentul verbal: dispoziții de a aproba sau respinge, de a susține sau a combate; sau pur și simplu de a discuta – de a folosi argumente pro și contra. Ca majoritatea celor care acceptă obiectele lumii a 2-a („mentalștii“), materialștii interpretează de regulă conținuturile lumii a 3-a ca și cum ar fi „idei în mintea noastră“:



materialiștii radicali încearcă însă mai mult, anume să interpreteze aceste „idei în mintea noastră” – și deci și obiectele lumii a 3-a – ca dispoziții ale creierului pentru comportamentul verbal.

Și totuși nici mentalistul nici materialistul nu pot să justifice în acest fel obiectele lumii a 3-a, în special conținuturile teoriilor și relațiile logice obiective dintre acestea.

Obiectele lumii a 3-a nu sunt nici „idei în mintea noastră”, nici dispoziții cerebrale pentru comportamentul verbal. Și nu schimbăm cu nimic situația dacă la aceste dispoziții adăugăm și întrupările lumii a 3-a menționate în primul paragraf al acestei secțiuni. Pentru că nici una din aceste concepții nu poate explica adecvat caracterul abstract al obiectelor lumii a 3-a și mai ales relațiile logice care există între ele <sup>3</sup>.

De pildă, *Grundgesetze* a lui Frege a fost scrisă și parțial publicată atunci când el a dedus, dintr-o scrisoare a lui Bertrand Russell, că există o contradicție în fundamentele teoriei sale. Această contradicție s-a aflat acolo, obiectiv, ani de zile. Frege n-a observat-o: ea n-a fost „în mintea lui”. Russell a semnalat-o (în legătură cu un manuscris diferit) într-un moment în care manuscrisul lui Frege era complet. Astfel, ani de zile a existat o teorie a lui Frege (și una similară, mai recentă, a lui Russell) care era în mod obiectiv inconsistentă fără ca cineva să aibă cea mai vagă bănuială despre acest lucru, sau fără ca starea creierului cuiva să-l determine să fie de acord cu afirmația „Acest manuscris conține o contradicție”.

Rezumând, putem spune că obiectele lumii a 3-a, proprietățile și relațiile lor nu pot fi reduse la obiecte ale lumii a 2-a. Și nu pot fi reduse nici la stări sau dispoziții ale creierului; nici chiar dacă am admite că toate stările și procesele mentale pot fi reduse la stări și procese ale creierului. Și asta în pofida faptului că putem considera lumea a 3-a ca fiind produsul minții umane.

Russell n-a inventat și n-a produs el însuși inconsistența, ci doar a *descoperit* că exista. (El a inventat sau a produs o modalitate de a arăta sau de a dovedi că inconsistența există). Dacă teoria lui Frege n-ar fi fost în mod obiectiv inconsistentă nu i s-ar mai fi aplicat dovada de inconsistență a lui Russell și

Frege însuși n-ar mai fi fost convins de această inconsistență a teoriei sale. Astfel, o stare a minții lui Frege (și fără îndoială o stare a creierului său) a rezultat, parțial, din faptul obiectiv că această teorie era inconsistentă: el a fost foarte afectat de descoperirea inconsistenței. Acest lucru l-a determinat să scrie apoi (un eveniment al lumii 1) următoarele: „*Die Arithmetik ist ins Schwanken geraten*“ („Aritmetica este nesigură“). Astfel, există o interacțiune între (1) evenimentul fizic sau parțial fizic al primirii de către Frege a scrisorii lui Russell; (2) faptul obiectiv, neobservat până atunci, aparținând lumii a 3-a, că există o inconsistență în teoria lui Frege; și (3) evenimentul fizic sau parțial fizic al scrierii de către Frege a comentariului său asupra statutului aritmeticii (lumea a 3-a).

Acestea sunt câteva dintre motivele pentru care susțin că lumea 1 nu este închisă cauzal și pentru care afirm că există o interacțiune (deși indirectă) între lumea 1 și lumea a 3-a. Și mi se pare clar că această interacțiune este mediată de evenimente mentale și parțial consistente ale lumii a 2-a.

Dar un fizicist nu poate admite, desigur, nici una din aceste afirmații.

Și cred că el se mai lovește și de altă problemă: nu poate justifica funcțiile superioare ale limbajului.

Această critică a fizicalismului se leagă de analiza funcțiilor limbajului care a fost introdusă de profesorul meu, Karl Bühler. El a distins trei funcții: (1) funcția expresivă; (2) funcția de semnalizare sau de transmitere și (3) funcția descriptivă. Am discutat teoria lui Bühler în diferite locuri și am adăugat la cele trei funcții ale sale o a patra funcție – (4) funcția argumentativă. Am argumentat în altă parte<sup>4</sup> că fizicistul poate explica numai primele două funcții. Ca rezultat, atunci când se confruntă cu funcția descriptivă și cu cea argumentativă ale limbajului, fizicistul vede întotdeauna numai primele două funcții (care sunt de asemenea prezente întotdeauna) și acest lucru are consecințe dezastruoase.

Pentru a vedea care este problema, este necesar să discutăm pe scurt teoria funcțiilor limbajului.

În analiza actului vorbirii, Bühler face distincții între *vorbitor* (sau, cum îl mai numește el, *expeditor*) și persoana căreia i se

vorbește, *ascultătorul* (sau *destinatarul*). În anumite cazuri speciale („degenerate“) destinatarul poate să lipsească sau poate să fie identic cu expeditorul. Aceste patru funcții (există și altele, cum ar fi funcția de comandă, de consiliere – de asemenea, enunțurile performative ale lui Austin <sup>5</sup>) se bazează pe existența unor relații între (1) expeditor, (2) destinatar, (3) alte obiecte sau stări de lucru care, în cazuri degenerate, pot fi identice cu (1) sau (2). Am să expun în continuare un tabel al acestor funcții, pe marginea căruia pot fi făcute următoarele observații:

(1) Funcția expresivă constă într-o exprimare exterioară a unei stări interioare. Chiar și instrumente simple cum ar fi un termometru sau lumina unui semafor își „exprimă“ stările în acest sens. Animalele (și uneori plantele) își exteriorizează stările interioare prin intermediul comportamentului. La fel procedează și oamenii. De fapt, orice activitate pe care o desfășurăm, nu numai folosirea limbajului, este o formă de auto-exprimare.

(2) Funcția de semnalizare (Bühler o mai numește și „funcția de transmitere“) presupune funcția expresivă și este deci la un nivel mai înalt. Termometrul ne poate semnala că este foarte frig. Semaforul este și el un instrument de semnalizare (deși poate continua să funcționeze chiar și atunci când nu circulă automobile pe stradă). Animalele, în special păsările, emit semnale de primejdie; chiar și plantele emit semnale (de pildă, pentru insecte); și atunci când comportamentul nostru (lingvistic sau nu) determină o reacție, la un animal sau la un om, putem spune că a fost luat ca un semnal.

(3) Funcția descriptivă a limbajului presupune cele două funcții anterioare. Caracteristica acestei funcții este aceea că, dincolo de exprimare și comunicare (care în anumite situații pot să fie aspecte lipsite de importanță) odată cu ea sunt introduse enunțurile *adevărate* sau *false*: adică sunt introduse standardele adevărului și falsității. (Putem distinge o parte inferioară a funcției descriptive unde descrierile false depășesc puterea de abstracție a animalelor (și a albinelor?). Aici poate intra și un termograf, pentru că descrie adevărul cu condiția să nu se deterioreze).

(4) Funcția argumentativă adaugă argumentul la celelalte trei funcții, cu valorile sale – valid / nevalid.

funcții		valori	} omul
animalele plantele	poate și albinele		
	(4) funcția argumentativă	valid / nevalid	
	(3) funcția descriptivă	adevăr / fals	
	(2) funcția de semnalizare	eficiență / ineficiență	
	(1) funcția expresivă	relevant / irelevant	

Funcțiile (1) și (2) sunt aproape întotdeauna prezente în limbajul uman; dar ele sunt, de regulă, mai puțin importante, cel puțin în comparație cu funcția descriptivă și cea argumentativă.

Și totuși, atunci când fizicalistul radical și behavioristul radical se ocupă de analiza limbajului uman, ei nu pot trece de primele două funcții <sup>7</sup>. Fizicalistul va încerca să dea o explicație fizică – o explicație causală – fenomenelor limbajului. Acest lucru este echivalent cu interpretarea limbajului ca o formă de exprimare a stării vorbitorului și deci cu reducerea lui la funcția expresivă. Behavioristul, pe de altă parte, va fi interesat și el de aspectul social al limbajului – dar numai din perspectiva felului în care acesta afectează comportamentul altor persoane; sub aspectul „comunicării”, ca să folosim un cuvânt la modă; ca modalitate în care vorbitorii își răspund reciproc la „comportamentele lor verbale”. Aceasta înseamnă că limbajul este redus doar la exprimare și comunicare.

Consecințele acestei perspective sunt însă dezastruoase. Dacă limbajul este redus numai la funcția expresivă și comunicativă, el pierde tocmai ceea ce îl deosebește de limbajul animal: abilitatea de a formula enunțuri adevărate sau false și de a produce argumente valide sau nevalide. Drept consecință, fizicalistul nu mai este capabil să facă diferența între propagandă, intimidare verbală și argumentare rațională.

### III. EPIFENOMENALISMUL

Dintr-un punct de vedere darwinist, putem să ne întrebăm, care este contribuția proceselor mentale la supraviețuirea speciei. De pildă, putem considera că durerea este un semnal de avertizare. La un nivel mai general, darwiniștii trebuie să privească „mintea“, adică procesele mentale și dispozițiile pentru acțiuni și reacțiuni mentale, ca fiind analoagă cu un organ corporal (legat strâns de creier, probabil) care s-a dezvoltat sub presiunea selecției naturale. Funcția sa ar fi aceea de a sprijini adaptarea organismului<sup>8</sup>. Perspectiva darwinistă este, pe scurt, următoarea: conștiința și procesele mentale în general, trebuie să fie privite (și, dacă este posibil, explicate) ca rezultat al evoluției prin selecție naturală.

Perspectiva darwinistă este necesară în special pentru înțelegerea proceselor mentale intelectuale. Acțiunile inteligente sunt acțiuni adaptate unor evenimente previzibile. Ele se bazează pe anticipare și expectație; de regulă, pe așteptarea de lungă și scurtă durată și pe compararea rezultatelor așteptate ale mai multor acțiuni și reacțiuni posibile. Aici intervine *preferința* și odată cu ea luarea deciziilor, dintre care o mare parte au o bază instinctuală. Aceasta este poate calea prin care emoțiile intră în lumea a 2-a, a experiențelor și proceselor mentale și motivul pentru care uneori ele „devin conștiente“, iar alteori nu.

Perspectiva darwinistă explică, de asemenea, cel puțin parțial, prima apariție a unei lumi a 3-a de produse ale minții umane: lumea uneltelor, instrumentelor, limbajelor, miturilor și teoriilor (Acest lucru poate fi admis și de către cei care șovăie sau ezită să atribuie „realitate“ unor entități cum sunt problemele și teoriile, ca și de cei care consideră că lumea a 3-a este o parte a lumii 2 și / sau a lumii a 2-a). Existența lumii culturale (lumea a 3-a) și a evoluției culturale, ne poate atrage atenția asupra faptului că există o coerență sistematică foarte mare atât în cadrul lumii a 2-a cât și în cadrul lumii a 3-a; acest lucru poate fi explicat – parțial – ca fiind rezultatul sistematic al presiunilor selecției. De pildă, evoluția limbajului poate fi explicată, se pare, numai dacă ne asumăm faptul că și un limbaj primitiv poate fi de folos în lupta pentru existență și ca apariția limbajului are un efect feedback: capaci-

tățile lingvistice sunt competitive; ele sunt selecționate pentru efectele lor biologice; ceea ce conduce apoi la niveluri tot mai înalte pe scara evoluției limbajului.

Putem rezuma toate acestea sub forma următoarelor patru principii dintre care primele două cred că trebuie să fie acceptate în special de cei care sunt înclinați spre fizicalism sau materialism.

(1) Teoria selecției naturale este singura teorie cunoscută până acum care poate explica apariția proceselor cu scop și în special a formelor de viață superioare.

(2) Selecția naturală se ocupă de supraviețuirea fizică (de frecvența distribuției genelor competitive într-o populație). Este deci legată în primul rând de explicarea efectelor lumii 1.

(3) Dacă teoria selecției naturale trebuie să explice apariția lumii a 2-a, lumea experimentelor mentale și subiective, atunci trebuie să explice și modul în care evoluția lumii a 2-a (și a lumii a 3-a) ne oferă în mod sistematic mijloace de supraviețuire.

(4) Orice explicație în termeni de selecție naturală este parțială și incompletă. Aceasta pentru că trebuie să asume întotdeauna existența mai multor mutații competitive și a unei variații de presiuni de selecție care sunt parțial necunoscute.

Aceste patru principii pot fi numite, pe scurt, punctul de vedere darwinist. Am să încerc să arăt că el este incompatibil cu concepția numită de obicei „epifenomenalism”.

Epifenomenalismul admite existența evenimentelor sau experiențelor mentale – adică existența lumii a 2-a – dar susține că aceste experiențe mentale sau subiective sunt produse secundare inefficiente cauzal ale proceselor fiziologice, singurele care au o eficiență cauzală. În acest fel epifenomenalismul poate accepta atât principiul fizicalist al compactității lumii 1, cât și existența lumii a 2-a. El trebuie doar să insiste că lumea a 2-a este irelevantă; că numai procesele fizice sunt importante. Dacă un om citește o carte, nu este decisiv faptul că acest lucru îi influențează opiniile și îi îmbogățește cunoștințele. Acestea nu sunt decât epifenomene irelevante. Ceea ce contează este numai schimbarea care se produce în structura creierului sau și care îi afectează dispoziția de a acționa. Aceste dispoziții, va spune epifenomenalistul, sunt de cea mai mare importanță pentru supraviețuire: numai aici intervine darwinismul. Experiențele

subiective, cum sunt cititul și gândirea, există într-adevăr, dar ele nu au rolul pe care li-l atribuim de obicei. Această atribuire greșită este un rezultat al faptului că nu reușim să facem distincția între experiențele noastre și impactul crucial pe care cititul, de pildă, îl are asupra proprietăților dispoziționale ale structurii creierului. Aspectele experiențiale subiective ale percepțiilor noastre în timp ce citim o carte sunt mai puțin importante; la fel și aspectele emoționale. Ele sunt fortuite, cazuale mai degrabă decât cauzale.

Este clar că această perspectivă epifenomenalistă nu este satisfăcătoare. Ea admite existența lumii a 2-a dar îi neagă orice funcție biologică. De aceea nu poate explica, în termenii darwiniști, evoluția lumii a 2-a. În plus, este silită să nege faptul cel mai important – impactul enorm al acestei evoluții (și al evoluției lumii a 3-a) asupra lumii 1.

Cred că acest argument este decisiv.

În termeni biologici, există în organismele superioare o seamă de sisteme de control strâns legate între ele: sistemul imunitar, sistemul endocrin, sistemul nervos central și ceea ce putem numi „sistemul mental”. Este aproape sigur că ultimele două sunt strâns legate între ele. La fel se poate spune și despre celelalte, chiar dacă legătura dintre ele este mai puțin strânsă. Sistemul mental are în mod sigur propria istorie funcțională și evoluționistă, funcțiile sale au sporit odată cu evoluția de la organismele inferioare la cele superioare. Prin urmare, el este inevitabil legat de punctul de vedere darwinist. Epifenomenalismul nu oferă însă nici o legătură.

## NOTE

<sup>1</sup> Vezi, de pildă, J. Huxley, *Evolution. The Modern Synthesis*, 1942; P.B. Medawar, *The Future of Man*, 1960; T. Dobzhansky, *Mankind Evolving*, 1962.

<sup>2</sup> Vezi Immanuel Kant, *Critique of Pure Reason*, ediția a doua, 1787.

<sup>3</sup> Pentru o discuție mai largă, vezi *The Self and its Brain*, partea I, secțiunea 21.

<sup>4</sup> Vezi, *Conjectures and Refutations*, capitolul 12.

<sup>5</sup> Vezi J.L. Austin, *How to Do Things with Words*, 1962.

<sup>6</sup> Albinele dansatoare transmit *poate* o informație factuală sau descriptivă. Un termometru sau un barometru face același lucru prin intermediul cifrelor. Este interesant că în ambele cazuri este exclusă problema minciunii – deși cel care fabrică termometrul ar putea să-l folosească pentru a ne da o informație greșită. (Pentru albinele dansatoare vezi K von Frisch, *Bees: their Vision, Chemical Senses.. and Language*, 1950; *The Dancing Bees*, 1955; și M. Lindauer, *Communication Among Social Bees*, 1961).

<sup>7</sup> Vezi nota 4.

<sup>8</sup> Vezi discuția despre evoluția organică din *The Self and Its Brain*, partea I, secțiunea 6.

Traducere Laurențiu STAIKU



# MINTE ȘI TRUP (*MIND-BODY*)\*

Kathleen V. WILKES

(Universitatea din Oxford)

## 1. INTRODUCERE

Cred că toți cititorii cunosc vechea și banala glumă: “*What is matter? Never mind? What is mind? No matter* (Ce este materia? Niciodată mind. Ce este mind? Nu materie)”. La fel de veche ca și această glumă este însă și interesanta problemă constând în așa numita dihotomie minte-trup (*mind-body*), problemă care a fost aproape sanctificată în filosofia post-carteziană și de care nu este deloc ușor să te aperi sau să o eviți. Aceasta înseamnă, desigur, că problema minte-trup este la fel de dificil de a fi descrisă și rezolvată, cât și, ceea ce mi se pare preferabil, de a fi înlăturată.

Vom începe prin a analiza conținutul distincției „mental/fizical”. Aceasta va constitui prima parte a articolului. Apoi, în a doua parte, voi încerca să ofer o scurtă descriere a principalelor premise ale înțelegerii relației dintre „mental și fizical”, sau dintre minte (*mind*) și creier (*brain*).

## 2. „MENTAL” ȘI „FIZICAL”

### a) Mentalul

Ar fi dificil de explicat termenul „mental” unui grec antic. La o lectură foarte grăbită a lui Aristotel, de exemplu, vom găsi

---

\* „Mind and Body”, în *Revue Roumaine de Philosophie*, 35/3–4/1991.

\*\* Joc de cuvinte, intraductibil în românește: Ce este materia? Nu contează? Ce este spiritul? N-are importanță.

termenul *psyché*, care în mod obișnuit, dar foarte greșit, este tradus prin suflet (și care este strămoșul termenului psihologie). Aristotelicul „*psychè*“ nu se reduce la acele fenomene pe care noi le desemnăm prin termeni ca „*minte*“ sau „*mental*“. Pentru Aristotel <sup>1</sup> tot ceea ce este viu are „*psychè*“ – tot atât arțarii, cât și țânțarii, șoarecii și oamenii. „*Psychè*“ include printre funcțiile sale vederea, gândirea, imaginația, dar și alte funcții ale viețuitoarelor ca digestia, creșterea, mișcarea, metabolismul.

„*Mentalul*“, așa cum îl concepem astăzi, își datorează sfera unor două surse foarte diferite. Prima este cartezianismul: intelectul este suma tuturor lucrurilor de care suntem direct și imediat conștienți. Astfel, reprezentarea mea actuală a unei cafele vieneze este *mentală* ca și durerea din picior sau amintirea unei furtuni la Sofia. Fenomenele mentale sunt înțelese ca fiind toate acele lucruri de care suntem imediat și incorigibil conștienți; noi avem un acces direct, nemijlocit și privilegiat la ele: „Conștiința nu înșală niciodată“ proclamă Hume <sup>2</sup>, urmându-l aici pe Descartes, așa cum o făcuseră înaintea sa și Locke și Berkeley. Astfel, de exemplu, o însușire ca percepția este inclusă *într-un anumit sens* în conștiință, dar nu ca „*percepție* a unui copac“, cât mai curând ca experiența, simțirea, senzația vederii unui copac, indiferent de ceea ce este comun între vederea actuală a unui copac și iluzia acestuia.

Această definiție a *mentalului*, care îl face co-extensiv cu intelectul conștient a fost revoluționară. Înaintea lui Descartes nimeni nu se îndoia că sunt mai multe acele lucruri despre psihicul uman pe care *nu* le cunoaștem decât acelea de care *nu* suntem conștienți. Cerința vechilor greci „*cunoaște-te pe tine însuși*“ ar fi fost lipsită de sens dacă faimoasa cunoaștere de sine ar fi fost atât de facilă. Heraclit spunea că nicicând nu vom putea descoperi profunzimile sufletului, oricât de departe și de insistent vom încerca. Scriitorii, poeții și filosofi mărturisesc cu toții, de secole, că suntem deosebit de confuzi, auto-amăgiți, ignoranți cu privire la noi înșine și la ceea ce se petrece în mintea noastră. Într-adevăr, este insuficient amintit faptul că primul scop al lui Descartes era de a combate scepticismul și *nu* de a oferi o descriere a intelectului (*mind*); interpretarea pe care el a dat-o lui *mens* a fost

estompată de strategia sa epistemologică. El gândea că singurul mod de a învinge scepticismul este de a găsi acele fundamente constante ale cunoașterii, acele fapte pe care nici cel mai aprig sceptic nu le-ar putea contesta. Descartes, credea că singurii candidați apti să joace rolul acestor fundamente de nezdruccinat sunt aserțiunile de tipul lui *cogito* – generalizate în raport cu orice conținuturi curente ale conștiinței lucide aflate la orice persoană întâi, timpul prezent. Aceste conținuturi trebuie să fie cât mai evidente, astfel de motivații epistemologice devenind teze metafizice sau ontologice: „gândirea nu este altceva decât conștiința lucidă“.

Evident aceasta nu se va realiza, deși ea face cu adevărat parte din poveste. Foarte cotidianul fapt al auto-decepției ne arată că ne putem înșela cu privire la noi înșine: nu orice este deschis privirii veșnic treze a Ochiului Tainic al introspecției conștiinței. Să analizăm, de asemenea, un alt fenomen cotidian: ați ajuns la capătul unei scări. Dacă vă împiedicați și cădeți veți spune: „Mă gândeam (sau, poate „Trebuie că mă gândeam“) că mai era o treaptă“. Dar ceea ce se afla în conștiința voastră lucidă în acel moment, erau, să presupunem, gânduri cu privire la sănătatea iubitului vostru canar. Pare adevărat că acest gând („încă o treaptă“) explică accidentul pe care l-ați avut, dar, orice ar reprezenta el, acest gând nu se afla în conștiința voastră lucidă. În plus, fie că suntem sau nu de acord cu Freud, cu toții admitem astăzi că factorii de care nu suntem conștienți constituie multe dintre ițele ascunse ale caracterului, personalității, motivațiilor noastre (Așa cum se vede, Freud nu a „descoperit“, inconștientul iar Descartes mai curând, l-a acoperit, echivalând „intelectul“ cu vârful celuiui iceberg care este „conștiința lucidă“ (*conscious mind*)).

Fără îndoială, Descartes a lăsat o moștenire bogată. Importanța ideii conform căreia noi avem un fel de acces special, privilegiat și incorigibil, cel puțin la unele dintre stările noastre mentale, este încă actuală și dă naștere celui mai tulburător aspect al problemei trup / suflet, așa cum vom vedea. Cu alte cuvinte, aș spune că „dilema trup / suflet“ este mult mai dificil de rezolvat atunci când este privită mai întâi ca problemă a raportului dintre fenomenele mentale *conștiente* și fenomenele fizice. Deocamdată această afirmație este doar un gaj care va fi achitat un pic mai târziu.

Înainte de a continua analiza noțiunii „carteziene“ a intelectului, vom examina a doua sursă a actualei sfere a termenului „mental“. Aceasta se află și mai în trecut, în scolastica medievală, dar a fost cu cea mai mare claritate formulată de Brentano (cu care, accidental, Freud a studiat un timp). Este ceea ce medievalii numeau „inexistență intențională“ (*intentional inexistence*); astăzi noi o numim „intensionalitate“ (*intensionality*) cu *s* și nu cu *t* la mijloc, pentru a o distinge de acea stare mentală specifică, starea de a intenționa să faci ceva.

În sens larg, o stare intensională este o stare care are conținut, care este *despre* ceva. Acest ceva poate să nu existe (altfel, puteți avea credința în Moș Crăciun, grecii credeau în Zeus), sau poate să nu fie adevărat (puteți crede că se află o pasăre dodo în baie). De asemenea, conținutul stării dumneavoastră mentale poate să existe sau să fie adevărat. Nu aceasta contează, ceea ce este adevărat este că aveți asemenea gânduri, dorințe, etc. indiferent dacă obiectul lor există sau nu. Un alt fel de a descrie intensionalitatea este următorul. Să presupunem că omul în haină maro este primarul Londrei. Voi aveți o anumită părere despre acel om – credeți că este agent KGB. Aceasta nu înseamnă că dumneavoastră credeți că primarul Londrei este agent sovietic (nu știți că omul în haină maro este primarul); de fapt, dumneavoastră știți că primarul este un fanatic anti-sovietic. Există o însușire pe care omul în haină maro o are – fiind bănuț de dumneavoastră a fi agent sovietic – dar pe care primarul nu o are, *chiar dacă* este una și aceeași persoană. Vom numi aceasta „absență a substitutivității“ („failure of substitutivity“), deoarece, dacă eu spun despre dumneavoastră că bănuț că omul în haină maro este spion KGB, eu nu pot înlocui expresia „omul în haină maro“ prin expresia „primarul Londrei“ atunci când relatez despre credința dumneavoastră. Absența pe scară largă a substitutivității nu înlocuiește însușiri non-mentale: dacă omul în haină maro are însușirea de a avea un pantof scâlțiat sau înălțimea de 1,80 m, atunci aceste însușiri le va avea și primarul. Vom reveni la aceasta; ea amintește de definiția leibniziană a identității: dacă *a* este identic cu *b*, atunci, desigur, tot ceea ce este adevărat despre *a* va fi adevărat și despre *b*. Fenomenele mentale par să încalce acest principiu: „a fi suspectat de Kathy Wilkes ca agent KGB“

este adevărat despre omul în haină maro, dar nu și despre primar. Brentano acordă intensionalității o imensă importanță, ca fiind cea mai profundă modalitate de a introduce o delimitare în sfera existenței, distingând mentalul de fizical.

Această distincție este, de asemenea, centrală pentru caracterizarea contemporană a mentalului. Ea este mai largă decât restrângerea carteziană la ceea ce este conștient, din moment ce include cu ușurință gânduri, dorințe, etc., non-conștiente. Pe de altă parte, însă, această distincție își are propriile neajunsuri; de fapt, acolo unde considerațiile „carteziene“ au succes, intensionalitatea eșuează adesea și invers. Se pare că intensionalitatea nu reușește să dea seama de unele fenomene mentale paradigmatic: „Mă doare“ nu se conformează pe deplin criteriului intensionalității dar este ușor de acomodat cu echivalența dintre mental și conștient. Invers, intensionalitatea pare să includă prea multe. Astfel, unii neuropsihologi descriu un anumit lob al creierului șobolanului ca pe un „comparator“ care compară „recompensele și pedepsele primite cu cele așteptate. „Așteptarea“ este o perfectă noțiune intensională, dar vrem noi cu adevărat să descriem informațiile din creierul unui șobolan drept stări mentale? Apoi, termenii intensionali sunt folosiți pentru a descrie, într-o manieră liberă, operații realizate de computere; dar este cel puțin o problemă deschisă: în ce măsură putem accepta ideea după care computerele au o viață mentală.

Ceea ce reținem este că nu există o caracterizare unică și unanim acceptată a ceea ce este *mental* și a ceea ce deosebește mentalul de non-mental sau fizical. Cele două alternative analizate sunt foarte diferite: prima este motivată epistemologic și se bazează pe o noțiune – „conștiință“ care nu a fost utilizată în limba engleză (sau franceză, sau germană) cu sensurile sale actuale până în a doua jumătate a secolului al XVII-lea. Cea de a doua constă dintr-un criteriu logic care se sprijină pe diferența dintre condițiile de adevăr ale enunțurilor cu privire la stări mentale și cele ale enunțurilor cu privire la stări fizicale. Noi le putem utiliza împreună, construind astfel un fel de noțiune hibridă care, pe de o parte, acoperă cele mai multe elemente pe care am vrea ca termenul „mental“ să le acopere și, pe de altă parte, nu include prea multe din cele pe care am vrea să le

excludă. Oricum, este posibil ca sursa problemei să conștie în aceea că nu există câtuși de puțin o graniță precisă și astfel „dihotomia“ analizată nu este nici limpede și nici prea importantă. Voi dezvolta această idee în a doua parte a articolului; pentru moment vreau doar să arăt că atunci când dorim să descriem alte persoane, noi folosim un mare număr de termeni a căror apartenență la sfera mentalului, respectiv „fizicului“, nu este evidentă: „are o durere de dinți“, „doarme“, „este obosit“, „este nervos“, „este un extrovertit“. De asemenea, există un mare număr de termeni „mentali“ care sunt folosiți în mod obișnuit de neuropsihiatri sau programatori pentru a descrie operații ale creierului sau programe.

## b) Fizicul

În timp ce dificultatea de a defini „mentalul“ este larg recunoscută, mulți consideră că nu există nici o dificultate în ceea ce privește „fizicul“. Altfel spus, filosofii adoptă o atitudine facilă, conform căreia „fizicul“ este identic cu „tot ceea ce reprezintă obiect al discursurilor savanților naturaliști“. Când este vorba despre raportul dintre gândire și creier, atunci fizicul constă în toate acele elemente despre care vorbesc neurofiziologii: neuroni, sinapse. Dar, adesea, o asemenea atitudine conduce la eliminarea a ceea ce este de o importanță crucială.

Există desigur celule nervoase în creier, miliarde de asemenea celule, dar fiecare celulă are o alcătuire internă deosebit de complexă și trebuie desfăcută în microconstituentele sale. Desigur, există un nivel chimic al descrierii fiecărui neuron dincolo de care există un nivel atomic, iar atomii au ei înșiși o înfricoșătoare de complexă alcătuire internă. În consecință, pur și simplu noi nu știm în ce măsură a explica operațiile gândirii ne obligă să examinăm operațiile *sub*-celulare ale creierului. Echivalent și reciproc: de ce să mergem la un nivel atât de „micro“ cum este neuronul. S-ar putea ca acest lucru să nu fie necesar: se pare că multe dintre explicațiile anumitor trăsături ale vieții noastre mentale derivă din operații ale unor *lanțuri* de

elule, sau ale unor *colecții* de asemenea lanțuri, sau din proprietăți ale unor rețele neuronale. Mai mult, neurologii vorbesc despre funcții realizate de componente „macro“-cerebrale: hipotalamus, lobul temporal, emisfera dreaptă.

Este un fapt evident, dar prea adesea ignorat de filosofi, că există numeroase nivele de descriere a fizicalului. (Filosofii adesea suprasimplifică aria a ceea ce cunosc prea puțin; este aici un comentariu simpatic al lui Austin<sup>3</sup>: „suprasimplificarea este ceea ce am putea numi ocupația preferată a filosofilor, dacă nu chiar singura lor ocupație“). Dacă ne referim doar la creier putem așeza la un capăt al spectrului „creierul“, iar la celălalt capăt quarc-uri și neutroni“. Orice alegere intermediară, cum ar fi „neuronul“ sau „emisfera stângă“ este arbitrară și necesită o justificare care adesea eșuează. Trebuie să abandonăm ideea că „fizicalul“ este ceva substanțial. „Creierul“ poate fi substanțial, dar „proprietățile manifeste ale excitației neurale“ reprezintă o noțiune *abstractă*, iar quarcurile nu au nimic substanțial. De asemenea, trebuie să ne eliberăm de ideea că există un nivel privilegiat de descriere a creierului la care „mentalul“ trebuie, într-un fel sau altul, raportat. Cred că este pe deplin adevărat că cea mai bună descriere a „fizicalului“ este tot ceea ce se înțelege din terminologia pe care savanții naturaliști o utilizează în teoriile lor. Dar atunci, trebuie să recunoaștem că există o mulțime ierarhică de nivele despre care neurologii vorbesc și că ei nu vorbesc doar despre entități substanțiale, ci și despre *funcții* fizicale. Aceste funcții nu includ doar procesele de la nivelul sinapselor, ci și „așteptările“ șobolanului sau capacitatea de a „recunoaște imagini“ a subsistemelor aparținând emisferei drepte.

### c) Distincția

Deci: care este valoarea dihotomiei mental-fizical? Sugestia mea este următoarea: „Mentalul“ (echivalent din perspectiva a ceea ce ne interesează cu „psihologicul“) este cel mai bine definit astfel: „tot ceea ce este selectat din terminologia științelor psihologice“. „Fizicalul“ – în măsura în care este implicată știința

creierului – va fi: „tot ceea ce este selectat din terminologia neurologiei“. Aceasta înseamnă că nu există nici o graniță între cele două realități, așa cum nu există nici o graniță dată de Dumnezeu între mușuroiul de pe pajiștea mea și Alpi. Există, de obicei, o diferență: neurofiziologii vorbesc rar despre nevroze, iar psihanaliștii au, în mod obișnuit, puține de spus despre neutron. Tot așa, eu pot să sar cu ușurință peste mușuroiul de pe pajiștea mea, dar prefer să-i las pe alții să se cațere pe vârfurile Alpilor. Dar, dacă nu există o graniță absolută, un punct de demarcație – așa cum nu există nici între deal și munte – trebuie să admitem că există o mulțime de termeni nedeterminați, aparținând în egală măsură „mentalului“ și „fizicalului“. Există astfel o ramură a științei pe care o numim „neuropsihologie“ (sau, ceea ce este același lucru: „psihofiziologie“). Această disciplină își are propriul aparat conceptual care, inevitabil încalcă așa-numita diviziune între neural și psihologic. Iată și un exemplu luat direct din literatura de specialitate: „sistemul septal-hipocampal trimite un semnal prin fornix către centrul de decizie din hipotalamusul medial“.

Lecția care trebuie reținută, o lecție mult deosebită, cred, de orice altă examinare a problemei gândire-creier, este aceea că nu trebuie să vorbim despre două entități net distincte. În schimb, avem de analizat o complexă și încâlcită ierarhie de interrelații gândire-creier.

### 3. RELAȚIILE GÂNDIRE-CREIER

#### a) Două surse principale ale complexității

Dacă cele arătate mai sus sunt corecte, atunci problema suflet-trup nu există, iar noi nu putem spera să găsim nici o relație între entități de tip mental și entități de tip fizical. Atât mentalul cât și fizicalul sunt extrem de eterogene. Puține lucruri sunt mai



puțin asemănătoare decât complexul lui Oedip și senzația unei înțepături sau percepția subliminală și o stare de iritabilitate. Stejarii și cuarcii sunt în egală măsură „fenomene fizicale“, ca și grămezile de aur și proprietățile rețelilor atomice. Această eterogenitate – vizând atât „mentalul“ cât și „fizicalul“ – conferă problemei noastre un caracter încălzit, dar are însușirea de a fi deosebit de plauzibilă.

O altă foarte importantă sursă a încurcăturii constă în problema dacă vrem să punem în relație tipuri de entități mentale cu tipuri de entități fizicale, sau simboluri cu simboluri, sau ceva intermediar. Pentru a înțelege importanța distincției să comparăm „toți bărbații sunt fii“ cu „bărbatul care ține pușca este fratele meu“. Toți bărbații, Adam fiind singura excepție, sunt identici cu fiul cuiva; tipul bărbat este în mod sistematic legat de clasa fiilor. Această relație este deci una tip-tip, în exemplul nostru una de identitate. Dar nu toți frații mei țin o pușcă și nu toți bărbații care țin o pușcă sunt frați cu mine; „bărbatul care ține o pușcă este fratele meu“ – fiind un exemplu de relație *simbol-simbol*, și în acest caz tot una de identitate. Astfel că atunci când analizăm relațiile mental-fizical trebuie să decidem dacă și când ne așteptăm la o relație mai tare tip-tip sau la forma mai slabă simbol-simbol. Este probabil adevărat că aproape toate felurile de durere fizică implică (printre altele) inflamarea fibrelor c din trunchiul cerebral. S-ar putea deci să existe o relație tip-tip între durerea fizică și inflamarea fibrelor c (chiar dacă nu o relație de identitate, din moment ce durerea implică și multe altele, nu doar inflamarea fibrelor c). Dar „gândul meu din 12.01 că mi s-a terminat cafeaua“ poate avea puține sau chiar nici o corelație sistematică cu ceea ce se petrece în capul vostru când aveți un asemenea gând, astfel spus, s-ar putea să nu existe nici un tip de stare cerebrală care să însoțească mai curând gândul – cafea, decât gândul – ceai, sau gândul – pastă de dinți. Acel gând din 12.01 ar putea fi identic cu un anume proces cerebral complex, dar într-o manieră simbol-simbol.

Data fiind această complexitate pe care ne-am asumat-o, să examinăm câteva teorii care candidează în competiția elucidării relației mental-fizical.

## b) Câteva teorii alternative

Cea mai simplă imagine asupra relației gândire-creier este cea a identității: durerea este chiar o anumită stare a creierului, gândul meu că trebuie să cumpăr o altă cafea, alta. Dar aceasta necesită o examinare prudentă.

(i) Dacă *a* este identic cu *b* atunci – întrucât identitatea este simetrică – *b* este identic cu *a*. O precizare banală poate, dar ea înseamnă că dacă toate evenimentele mentale sunt fizicale, atunci unele elemente fizicale sunt mentale; relația de identitate nu are favoriți. Desigur, anumiți autori, care se auto-denumesc „teoreticieni ai identității”, vor ca fizicalul să fie într-un fel mai important, mai fundamental decât mentalul. Vom reveni imediat. O a doua complicație a oricărei teorii a identității este cea care rezultă din legea lui Leibniz, mai sus menționată: toate însușirile (atât cele intensionale, cât și altele) adevărate pentru *a* trebuie să fie adevărate și pentru *b*. Astfel dacă imaginea mea mentală este cea a siluetei unui cal, sau dacă durerea mea este cea de înțepătură ele nu pot fi identice cu nici o stare cerebrală, din moment ce acestora le lipsesc asemenea însușiri. Dar această dificultate poate fi evitată printr-o scamatorie: putem decide să identificăm *nu* imaginile sau durerile, ci starea de *a* avea imagini sau dureri cu anumite stări cerebrale. Desigur „a avea imaginea unui cal” sau – ceea ce este echivalent – „a vizualiza un cal” nu este totuna cu ceea ce am putea numi „formă cabalină” iar „a avea o durere” nu este identic cu „a fi împuns”. O a treia complicație este dacă ne așteptăm la identități de tipuri, sau la identități de simboluri sau la identități mixte tip-simbol.

Cum am putea proba sau imagina ipoteza unor asemenea identități? Un test empiric ar fi neconcludent dintr-un motiv foarte simplu: orice evidență în favoarea identității este într-o măsură egală o evidență în favoarea paralelismului dualist. Dacă un om de știință stabilește faptul că oricând cineva simte o durere aceasta este legată de o secvență de procese cerebrale oarecare, adeptul dualismului poate cu ușurință să replice că aceasta arată doar că oricând există durere există și secvența respectivă; el nu poate fi forțat să admită că ele sunt identice. „Oricând A, apoi B”

este la o distanță respectabilă de „A este B“. Oricând există o inimă atunci acolo este și un rinichi, dar inimile nu sunt rinichi.

Parțial din acest motiv, cel mai dezbătut și celebru argument în favoarea identității este unul care nu se bazează pe evidența empirică, ci încearcă o demonstrație *a priori*: este unul din argumentele lui Donald Davidson. El se raportează la trei premise. Prima, mentalul și fizicalul interacționează cauzal. A doua: oriunde există cauzalitate există și o anumită lege – orice stări cauzale singulare reale presupun o *anume* lege generală. A treia: nu există legi care să relaționeze fenomene mentale cu orice altceva, mental sau fizical; orice lege trebuie să relaționeze doar evenimente descrise cu termeni fizicali. Ei bine, nu este aici momentul să evaluăm cele trei premise; cel puțin a treia necesită un mare efort pentru a fi examinată și apărată. Dar dacă premisele sunt acceptate, atunci concluziile decurg astfel: să presupunem că un eveniment fizical – sirena unei ambulanțe – produce un fapt mental – dumneavoastră auziți zgomotul. Dacă primul fapt îl cauzează pe al doilea, atunci, aici trebuie să existe o lege (orice stare cauzală singulară presupune o anumită lege). Dar nici o lege nu relaționează fenomene mentale (senzația de zgomot) cu nimic altceva. Toate legile sunt de tip fizical-fizical. Deci, faptul mental al senzației de zgomot trebuie să fie simbolic – identic cu un fapt fizical. (Rețineți, doar o identitate simbolică: dacă nici o lege nu relaționează mentalul cu fizicalul atunci nu pot exista corelații tip-tip între acestea). Pentru că acesta este un aspect important îl voi întâri analizând un caz analog. Să presupunem că spuneți: „faptul relatat în pagina 3 din *The Times* a cauzat faptul relatat în pagina 4, din *The Morning Star*“. Să presupunem apoi că primul fapt a fost un cutremur de pământ într-o zonă populată, iar al doilea, distrugeri de bunuri și vieți omenești. Desigur, nu există legi care să reunească faptele descrise în pagina 3 din *The Times* cu faptele descrise în pagina 4 din *The Morning Star*. Murdoch nu a preluat încă și *The Morning Star*. Oricum, un fapt constă dintr-un cutremur, celălalt dintr-o distrugere de bunuri; dacă există o judecată adecvată a condițiilor date atunci vom ajunge la o lege care leagă cutremurele de o anumite forță de distrugere a clădirilor aflate într-o anumită stare în zona aceluia cutremur.

Din nefericire problema nu poate fi așa grăbit rezolvată din moment ce nu oricine poate accepta premisele lui Davidson. Mai mult, tezele sale nu trebuie considerate teze despre identitate. Ele ne oferă o concluzie monistă: nu există două feluri de lucruri sau stări, ci doar unul.

(ii) *Monism fără identitate (A): materialism eliminativ.* Oricum, am putea prefera să gândim fenomenele mentale ele nu sunt *identice* cu fenomenele fizicale, ci mai curând sunt *făcute din, constituite din nimic altceva decât din fenomene fizice*, la fel cum o masă este făcută din, constituită din nimic altceva decât molecule. Să reținem că relația „*x* este făcută din *y*” nu este simetrică: moleculele nu sunt făcute din mese. Prin urmare, nu este o relație de strictă identitate. Relația dintre mental și fizical este o relație pe care am putea-o exprima astfel: „*x* nu este altceva decât *y*”; de exemplu: „durerea nu este altceva decât inflamarea fibrelor *c*”. Această relație nu este simetrică, deci nu trebuie să acceptăm și reciproca.

Această poziție generează o versiune moderată și una extremistă. Aderenții moderați consideră că afirmațiile conform cărora „macrofenomenele” (mese sau dureri) sunt descrise ca fiind „făcute din” sau „nimic altceva decât” microfenomene (aglomerări de molecule sau procese cerebrale) sunt astfel susținute și validate: există realmente astfel de lucruri ca mese sau dureri, iar această analiză ne arată exact ceea ce ele sunt. Aderenții extremiști ai acestei interpretări vor susține că relația „nimic altceva decât” permite următoarea afirmație: „deci *x* nu există cu adevărat”, în felul în care afirmația „vrăjitoarele nu sunt nimic altceva decât femei nevrotice” permite concluzia: „deci vrăjitoarele nu există cu adevărat”. Imaginea pe care am oferit-o este deci aceea că, într-un viitor „splendid și imprevizibil”, nu vom mai auzi copii strigând „Mamă, mamă, mă doare piciorul”, ci copii care vor striga: „Mamă, mamă, sunt în starea cerebrală B.531.c”.

Aș fi dispusă să le răspund extremiștilor: ei bine, poate: vom trăi și vom vedea. Mai serios vorbind, s-ar putea să existe anumite fenomene mentale pe care într-o zi le vom accepta ca fiind „nimic altceva decât stări sau procese cerebrale; mai mult, s-ar putea ca în viitor să abandonăm descriția „mentală” a acestor fenomene, utilizând în schimb una fizicală. Oricum, pentru ca aceasta să se întâmple, trebuie să descoperim relații *tip-tip* între „vechile”

fenomene mentale și „noua“ descriere fiziologică; altfel, nu vom putea elimina cu ușurință vechea terminologie: s-ar putea să fie „nimic altceva decât“ o stare cerebrală diferită ori de câte ori același fenomen mental se produce. Pentru marea majoritate a lucrurilor despre care vorbim, și trebuie să vorbim, materialismul eliminativ extremist este foarte plauzibil. „Durere“ și „masă“ sunt cuvinte simple, scurte, ușoare. În schimb, seriile de evenimente în și prin care creierul este implicat în experiențele de durere și aglomerările de molecule din care este făcută o masă sunt incredibil de complexe și ar lua mult timp pentru a fi descrise. Acest aspect apare mai limpede dacă vom lua un exemplu diferit: este ridicol să ne imaginăm că am putea găsi vreodată un proces cerebral care să se coreleze neproblematic și sistematic cu ideea mea, a dumneavoastră și a oricui altcineva, că lampa cu gaz este mai romantică decât cea electrică.

*Paranteză: nu este nimic jignitor la adresa psihologiei în cele arătate.* Nu este nimic ciudat sau caraghios în faptul brut care ne spune că unele fenomene mentale au, în cel mai bun caz, doar explicații de tip simbol-simbol în neurofiziologie. Acest lucru este la fel de adevărat și pentru lumea fizică. Este clar că nu există nici un fel de corelații sistematice între garduri, scrumiere și ornamente, pe de o parte și microstructura lor, pe de altă parte. Științele – psihologia și fiziologia inclusiv – cercetează mai întâi obiecte naturale – clase de obiecte care sunt guvernate de legi sistematice cum sunt aurul, apa, tigrii și nu scrumiere, garduri, uniforme de poliție. Astfel, relații sistematice psiho-fizice vor apare doar între ceea ce aceste clase mentale și fizice se dovedesc a fi; este mai posibil că asemenea categorii vor include mai curând „recunoașterea imaginii“ decât „opinia că lămpile cu gaz sunt mai romantice decât cele electrice“.

(iii) *Monism fără identitate (B): funcționalism.* Până acum am schițat printr-un număr de enunțuri ce este *monismul*: presuposiția că fenomenele mentale nu sunt nimic altceva decât fenomene fizice. Teoria identității nu are favoriți; teoriile de tipul: „compus din“, „nimic altceva decât“ acordă preferință și prioritate fizicalului. Să ne întoarcem acum spre o abordare diferită a (unei părți a) problemei relației gândire-creier. Aceasta este perspectiva numită funcționalism pe care o vom putea analiza mai bine

dacă vom lua mai întâi în considerare behaviorismul clasic, din moment ce funcționalismul poate fi privit ca un „behaviorism reiterat“.

Behavioriștii au hotărât să interzică orice discuție despre mental (unii au negat că fenomenele mentale există, alții au considerat că referirile la ele nu sunt necesare). Organismul este o mare cutie neagră, iar comportamentul (*output-ul*) este o funcție a stării sistemului, și a stimulului (*input-ul*). Orice comportament poate fi în principiu explicat ca o funcție a acestor două variabile.

Ei bine, aceasta a fost o manieră prea simplistă de lucru, chiar dacă a putut avea o lungă carieră. În cele din urmă ea s-a dovedit incapabilă să explice comportamentul unui șoarece, cu atât mai puțin al unui singur om. Într-un fel, funcționalismul a cules ideea esențială și a dus-o mai departe. El nu s-a mulțumit cu singura mare cutie a behaviorismului, ci a descris-o și a introdus în ea o serie de alte cutii. Fiecare dintre aceste cutii mai mici are propriul său „comportament“, *output-ul* său este funcție de starea sa dată și de *input* (care poate fi din lumea externă sau de la o altă cutie; cutiile / funcții ulterioare se conectează, sau se alimentează, sau se sprijină pe cutiile / funcții anterioare).

Deci, fiecare cutie își are propriul *input* și *output*: *input-ul* poate proveni din lumea externă (de exemplu, cutia  $f_1$  din diagramă care primește informații din afară și le transmite ca *inputuri* (către alte funcții); alte cutii pot trimite semnale spre „cutia *output*“ ( $f_3$  în diagramă) care, să spunem activează mușchii pentru a produce sau opri o mișcare (mersul, vorbitul). Cele de la mijloc primesc *input-uri* de la și trimit *output-uri* spre celelalte funcții ( $f_2 - f_x, f_y$  în diagramă).

Ceea ce am arătat este foarte abstract și ar putea la fel de bine să descrie o fabrică de tractoare sau un sistem bancar, ca și un organism inteligent. Dar noi putem „îmbrăca“ acest model formal notând că aceste „cutii“, aceste funcții pot fi echivalente cu fenomene psihologice. Ele ar putea fi *teorii* pentru explicarea comportamentului. Astfel, o cutie poate fi etichetată „dorință“, alta „credință“, alta „memorie“, alta „percepție“. Sau, pentru a lua un exemplu actual din neuropsihologie, etichetele ar putea fi: „clasificator“, „generator de așteptări“, „comparator“, „amplificator“. Aceste etichete indică funcții mentale, iar schema

încearcă să arate cum acestea interacționează pentru a explica comportamentul. De exemplu, dacă vedeți o viespe pe mâna dumneavoastră veți dori să scuturați mâna pentru a o goni datorită credinței că viespea ar putea să vă înțepe și dorinței de a evita durerea. Dar dacă, în plus, știți că vă aflați pe marginea unei prăpastii, atunci acea informație, împreună cu informația memorizată despre efectele unei asemenea mișcări în situația unui echilibru precar poate avea ca efect suprimarea mișcării.

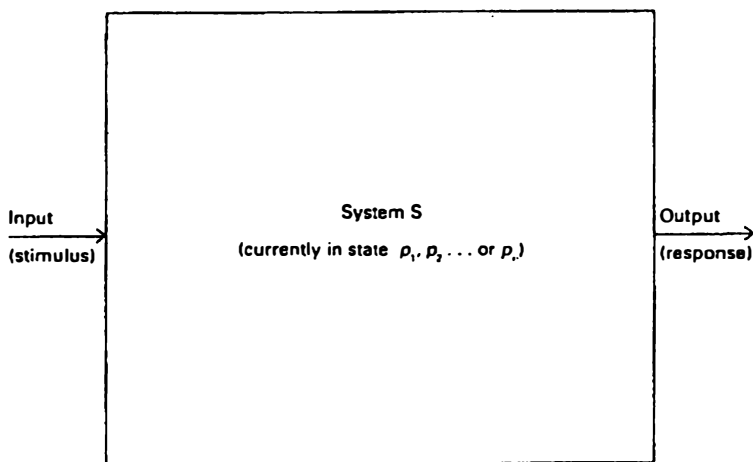


Fig. 1

Metoda poate fi, desigur, reiterată. Să presupunem o cutie, o funcție numită „memorie“. Ei bine, să presupunem că putem deschide această cutie și introduce o serie de cutii care contribuie la funcția dată: memoria de scurtă durată, memoria de lungă durată, memoria mediului, a propozițiilor, memoria iconică. Înseamnă că am creat o serie de sub-funcții (etichetate „ $s-f$ “ în diagramă). Apoi: să considerăm doar memoria de scurtă durată și să presupunem că ea este cutia centrală în diagrama 3, cutie pe care am numit-o „sub-funcția  $sfm$ “. Cum se realizează această

funcție? Am putea avea nevoie de sub-sub-funcții ca „mecanisme de re-aducere a informației” (numite *s-s-f* în diagramă). Iarăși o nouă serie subordonată de cutii. Ideea de bază este că, în principiu, cineva ar putea continua multă vreme să desfacă funcțiile psihologice complexe în comportamente tot mai simple până când – în principiu – ajungem la funcții suficient de simple

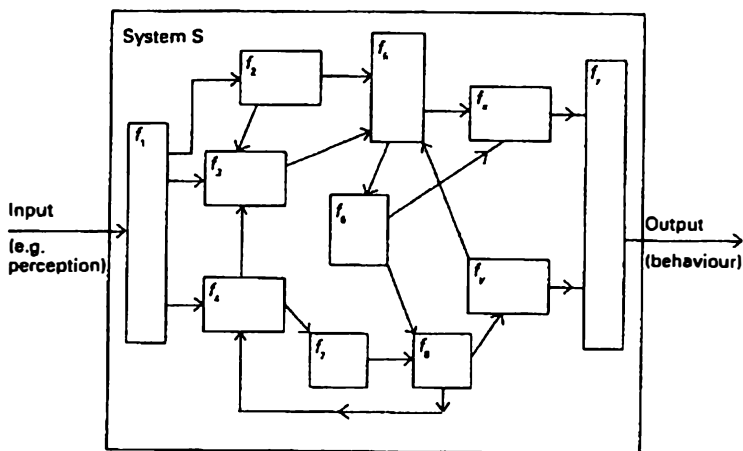


Fig. 2

ca „percepția limitelor” pentru care am putea căuta corespondențe neurofiziologice.

Această perspectivă oferă o informație suplimentară asupra problemei identității. Într-adevăr, ea ne spune că: „identitatea” este o categorie falsă: nu pot fi echivalate funcțiile cu lucrurile care realizează acele funcții, așa cum nu putem identifica deschiderea sticlei cu tirbușonul. Deschiderea sticlei este ceea ce tirbușonul face, nu ceea ce tirbușonul este. Similar, „detectia limitelor” este ceea ce fac celelalte tactici, nu ceea ce ele sunt. Dar aceasta nu este o poziție dualistă. Unul și același lucru poate fi descris sau *prin* funcțiile sale, sau *prin* structura sa; astfel un grup de celule poate fi numit „detector de margini”, așa cum tirbușonul poate fi numit „deschizător de sticle”. Sau, să privim lucrurile altfel: Distincția funcție/structură poate fi – și adesea este – comparată



în cazul unui computer cu distincția dintre software și hardware, dar această distincție nu este absolută, ci relativă față de nivelul la care vă raportați la computerul dumneavoastră. Scriind în limbajul cod-mașină cineva scrie un program software; la un nivel mai înalt avem limbajul de asamblare, care este un alt program, dar în raport cu care limbajul cod-mașină apare ca echipamentul, hardware-ul, cu care se lucrează. Mai sus, lucrând cu ADA, PASCAL

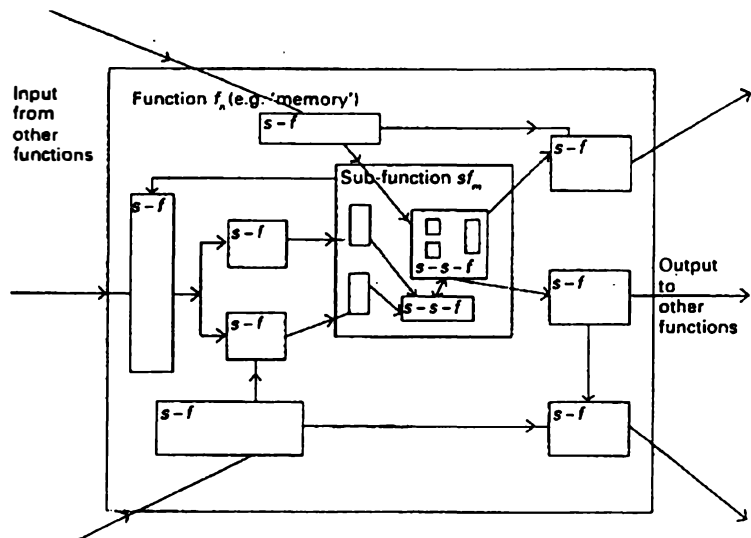


Fig. 3

sau BASIC programamtorul le tratează pe cele anterioare ca echipamente.

A caracteriza funcționalismul constituie o adevărată scamatorie pentru că el este de prea multe feluri. Sunteți acum în măsură să bănuieți că apare o mare diferență între versiunea sa științifică și cea corespunzătoare *simțului comun*: depinzând de faptul că funcțiile, cutiile selectate sunt mai curând asemenea apei sau aurului – aflate în relații sistematice (tip-tip) cu mecanismele fiziologice care le explică în ultimă instanță, sau sunt mai curând asemenea scrumierelor și gardurilor pentru care avem în cel mai bun caz relații simbol-simbol. Pe de altă parte, savantul neurolog rus Sokolov<sup>5</sup> explică obișnuințele la șobolani, postulând existența

unor funcții pe care le-a numit: „generatoare de așteptări“, „comparatoare“, „amplificatoare“ – aceștia nefiind termeni „mentali“ – tipici, așa cum îi gândim noi. Dar filosofii, care de regulă sunt tentați să examineze psihologia simțului comun, pot folosi terminologia cotidiană, familiară a „credințelor“, „amintirilor“, „dorințelor“.

(iv) *Probleme pentru funcționalism*. Se poate spune că funcționalismul reprezintă un pas înainte în problema minte-trup; el afirmă despre creier că este capabil (într-un fel oarecare) să producă software-ul stărilor mentale. Sper însă că nu veți fi surprinși să aflați că funcționalismul – indiferent de forma pe care o îmbracă – nu este un panaceu universal. Mulți autori cred că în acest fel „pot fi mânuite multe noțiuni intensionale de tip „Brentano“ – gândurile noastre, dorințele, credințele, amintirile. Acestea sunt însușirile psihologice pentru care – la fiecare nivel de descriere – structurile cerebrale sunt ultimele responsabile. Dar multe aspecte rămân neexplicate. În particular, funcționalismul pare să eludeze la fel ca și „mentalul“ cartezian: „durerile noastre, experiențele vizuale și celelalte. Pentru a pune problema mai clar să presupunem un computer care are toate „cutiile“ pe care bănuim că le avem în cap pentru a explica experiența vizuală. Deci, la fel ca și noi, el poate să spună când ceva este roșu sau albastru și așa mai departe. Dar o întrebare rămâne: are computerul *experiența calității de a fi roșu*? El poate discrimina, dar există aici, ca să folosesc celebra întrebare a lui Tom Nagel, ceva care „este ca și“ vizualitatea culorii roșu? Sau, pentru a pune această problemă generală într-un alt fel, mai direct: cunoștințele de robotică ne permit să concepem un robot care să aibă ceva similar fibrelor c, care să poată fi „stimulate“ cu o „informație stocată“, astfel încât să poată evita acele situații care i-ar pune în primejdie integritatea fizică; un robot sare și țipă când este zgâlțâit, are săculeți cu apă care să-i alunece din „ochi“ pe obrajii săi humanoizi, dar nimic din comportamentul său nu dă înțelesul de durere. Nagel crede că există „ceva care este ca și un om sau un liliac, dar nimic care să fie ca și un computer, așa cum nu există nimic care să fie ca un hamburger“.

Poate avem nevoie de o mică porție din ambele: o mică porție din discuția „clasică“ asupra identității și una din funcționalism.

Aceasta ar putea fi într-un deplin acord cu primele mențiuni conform cărora mentalul și fizicalul sunt *eterogene*: de ce să ne așteptăm să existe o singură strategie de alăturare a lor? Am putea specula, de exemplu, că procesele vederii necesită o interpretare funcționalistă a diferitelor funcții subordonate, care să explice abilitatea noastră perceptivă; aceasta ar putea explica cu succes capacitatea noastră de a discrimina culori și forme. Dar aspectul subiectiv sau „experimental“ al lucrurilor ne poate obliga să identificăm *experiența vizuală* cu ceva ca „frecvența înaltă a relației neurale“. Poate în prezent noi nu putem înțelege cum ceva ca *senzația de a vedea roșu* poate fi identificată cu o „frecvență înaltă“. Dar atunci cum se face că ne-fizicienii înțeleg identificarea materiei cu energia? Ne-fizicienii tind să o accepte – deoarece *Scientific American* sau suplimentul său duminical ilustrat le-o spune – chiar dacă, în felul ei, identitatea materie-energie este pentru profani tot atât de stranie.

#### 4. CONCLUZIE

Problema relației minte-corp trebuie mai curând clarificată decât rezolvată. Cu cât aflăm mai multe despre felul în care creierul este organizat spre a susține însușirile noastre mentale, cu atât vom fi mai puțin înclinați să simplificăm relația mental-fizical. De exemplu, ceva ca „detecția limitărilor“ este într-un fel ceva „mental“, dar neurofiziologii au descoperit deja multe în legătură cu felul în care oamenii – sau șoarecii, sau țânțarii – detectează marginile și, astfel, tot mai puțini cred că am putea avea o problemă filosofică profundă în legătură cu această subfuncție psihologică legică și non-conștientă și cu relația sa față de fizical. Altfel spus, computerele pot fi programate să detecteze contururi. Stările psihologice mai „sexy“, mai impresionante, ca sentimentul Weltanschauung-ului, al euforiei, al bucuriei transcendente sunt mult mai departe de orice explicație neuro-scientistă pentru că sunt mult mai complicate. Aspectele conștiente,

subiective, experiențiale ale vieții noastre mentale vor părea întotdeauna mai problematice, deoarece ne este foarte greu să acceptăm ideea că ele pot fi în principiu disecate, desfăcute în operații automate ale neuronilor și sinapselor. Oricum, rămâne adevărat faptul că un enunț cum ar fi: „«Cina cea de taină» a lui Leonardo da Vinci nu este altceva decât o colecție de molecule de vopsea pe o pânză” constituie o pierdere a tot ceea ce este, cel puțin virtual, semnificant.

## NOTE

<sup>1</sup> Aristotel, *De Anima*, cartea II, III, Oxford, Clarendon Press, ediția 1968.

<sup>2</sup> D. Hume, *An Enquiry Concerning Human Understanding*, Oxford. Clarendon Press, 1963, cap. VIII, partea I, § 52, p. 66.

<sup>3</sup> J.L. Austin, *How to Do Things with Words*, Oxford; Clarendon Press, 1962.

<sup>4</sup> D. Davidson, „Mental Events“ în *Experience and Theory* (ed. L. Foster și L.W. Swanson), University of Massachusetts Press, 1970, p. 79–101.

<sup>5</sup> Pt. teoria lui Sokolov, vezi J.A. Gray, *Elements of a Two-Process Theory of Learning*, Londra: Academic Press, 1975, p. 22.

<sup>6</sup> T. Nagel, *Mortal Questions*, Cambridge University Press, 1979, cap. 12 „What is it Like to be a Bat?”

Traducere Maria VORNICU și Angela BOTEZ

# EXPERIMENTELE MENTALE: O IPOTEZĂ PLATONISTĂ \*

James R. BROWN

(Universitatea din Toronto)

Metaforele despre cartea naturii scrisă în limbajul matematicii sugerează o legătură strânsă între matematică și fizică. Dar puțini filozofi contemporani, dacă există vreunii, consideră aceasta drept opinia lor fermă. Bănuiesc că marea majoritate susține o concepție cam de acest fel: matematica este o teorie a sistemelor formale; ea nu are nimic de a face cu lumea fizică, deși problemele din științele fizice pot sugera piste pentru cercetarea matematică. Rezultatele matematicii sunt în întregime cunoscute a priori prin derivarea lor din axiome luate ca adevărate prin convenție sau în virtutea înțeleșului termenilor cuprinși în acestea.

În opoziție, imaginea obișnuită a fizicii și a altor științe ale naturii este aproximativ aceasta: Fizica este o încercare de a descrie lumea fizică. Matematica furnizează instrumente utile, în sensul că teoretizarea presupune folosirea metodelor matematice, iar aceste modele matematice sunt date a priori. Dar cunoașterea noastră a faptului că lumea fizică este în acord cu unele structuri matematice particulare este, ea însăși, a posteriori și deci, spre deosebire de matematică, este failibilă.<sup>1</sup>

În bună măsură, această idee este rezumată într-o remarcă des citată a lui Einstein: „... în măsura în care propozițiile matematicii se referă la realitate, ele nu sunt certe; și în măsura în care ele sunt certe, nu se referă la realitate”<sup>2</sup>. Einstein continuă prin a explica de ce susține acest punct de vedere. Bănuiesc că motivele sale sunt îndeajuns de universale.

---

\* *Thought Experiments: a Platonic Account*. O parte a acestui eseu este extrasă din cartea mea, în curs de apariție, despre acest subiect. Pentru o discuție completă despre experimentele mentale, cu o mulțime de exemple, vezi *The Laboratory of the Mind: Thought Experiments in the Natural Sciences*, Routledge, 1991.

„... claritatea deplină asupra acestei stări de lucruri devine un loc comun numai prin acea tendință în matematică, cunoscută sub numele de „axiomatică“. Progresele obținute cu ajutorul axiomaticei, constau în aceea că s-a separat net forma logică de conținutul său obiectiv sau intuitiv; conform axiomaticei, numai formalul logic alcătuiește obiectul matematicii, obiect care nu are nimic comun cu conținutul intuitiv sau de alt gen asociat cu formalul logic.

Pe scurt, metodele și ontologia matematicii sunt în întregime diferite de metodele și ontologia fizicii; fizica este empirică și este despre lumea materială, în timp ce matematica nu este nici una, nici alta. Și mai scurt: matematica și fizica sunt extrem de diferite.

## INTUIȚIA MATEMATICĂ

Într-un articol foarte influent în filosofia matematicii, Paul Benacerraf<sup>3</sup> formulează o dihotomie și o dilemă. El remarcă două abordări principale ale matematicii: cea combinatorială și cea semantică. Concepția schițată mai sus este combinatorială, în timp ce platonismul este exemplul paradigmatic al unei abordări semantice. Dilema lui Benacerraf este aceasta: virtuțile unei abordări semantice sunt păcatele celeilalte. De exemplu, platonismul este foarte bun în a face inteligibilă ideea de adevăr matematic – propozițiile matematice sunt adevărate în același mod în care sunt adevărate propozițiile fizicii. În opoziție, orice gen de abordare combinatorială va avea de conexat adevărul cu demonstrabilitatea, ceea ce se va dovedi a fi extrem de problematic dintr-o mulțime de motive. (De exemplu, teorema lui Gödel arată foarte clar că adevărul și demonstrația sunt noțiuni distincte). Pe de altă parte, abordarea combinatorială face dreptate cunoașterii noastre – noi știm ceea ce știm deoarece avem o demonstrație. Platonismul, dimpotrivă, pare să se folosească de unele procese foarte misterioase pentru a da seama de modul cum

noi, oamenii, care există în spațiu și timp, putem ajunge să cunoaștem câte ceva despre entități abstracte în afară de spațiu și timp. Deci, putem avea o epistemologie acceptabilă sau o semantică acceptabilă, potrivit lui Benacerraf, dar nu și ambele.

Prin „demonstrație“, în abordarea sa combinatorială, Benacerraf înțelege o derivare. Adică, avem o demonstrație a lui  $P$  atunci când derivăm  $P$  dintr-o mulțime de axiome acceptate. Deci, atunci când demonstrăm  $P$ , arătăm, de fapt, că „ $\vdash \text{Axiome} \supseteq P$ “ este un adevăr logic. Nu mă îndoiesc nici un moment că o asemenea relație logică are loc, dar putea avea realmente dreptate Benacerraf, atunci când considera că această relație de derivare este temeiul pentru *cunoașterea* noastră că  $P$ ?

Să considerăm acest exemplu simplu. Ni se dă o tablă de șah cu două colțuri opuse înlăturate. Ni se dă de asemenea, un număr nelimitat de piese de domino, fiecare putând acoperi două pătrate ale tablei. Problemă: poate fi acoperită tabla de șah cu piese de domino, astfel încât fiecare pătrat să fie acoperit și să nu existe nici o suprapunere? Răspunsul este nu. Voi da două demonstrații.

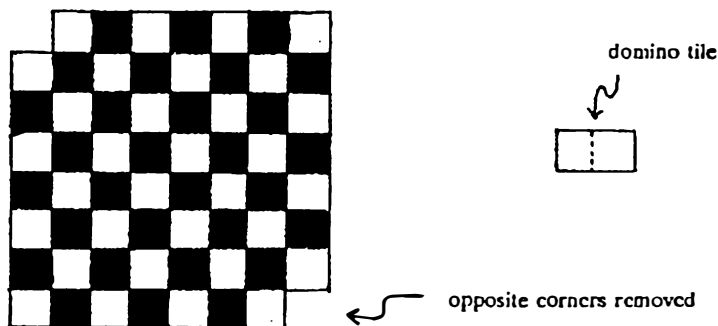


Fig. 1

Prima demonstrație nu este în nici un fel combinatorială. Se desfășoară după cum urmează: tabla nu poate fi acoperită, întrucât fiecare piesă acoperă un pătrat colorat și unul necolorat. Astfel, o acoperire completă cere un număr egal de pătrate colorate și necolorate, ceea ce nu avem pe această tablă. Deci, acoperirea tablei este imposibilă.

Iată o a doua demonstrație, care este combinatorială. Se etichetează fiecare pătrat cu literele  $P, Q, R$  etc., însemnând „ $P$  este acoperit“, „ $Q$  este acoperit“ ș.a.m.d.

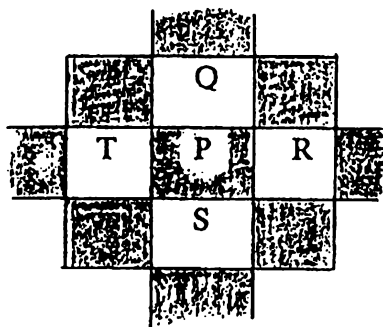


Fig. 2

O piatră care acoperă  $P$ , acoperă, de asemenea, un pătrat și numai unul în afară de  $P$ . O descriere completă a acoperirii din jurul lui  $P$  este:

$$\begin{aligned} &P \& Q \vee P \& R \vee P \& S \vee P \& T \& \\ &P \& Q \rightarrow \sim (P \& R \vee P \& S \vee P \& T) \& \\ &P \& R \rightarrow \sim (P \& Q \vee P \& S \vee P \& T) \& \\ &P \& S \rightarrow \sim (P \& Q \vee P \& R \vee P \& T) \& \\ &P \& T \rightarrow \sim (P \& Q \vee P \& R \vee P \& S) \end{aligned}$$

Aceasta este descrierea *numai pentru un pătrat*. Descrierea completă pentru întreaga tablă de șah va fi de șaiszeci și patru de ori mai lungă. Conjuncția tuturor acestor propoziții este o contradicție. Astfel, negația sa este o tautologie și, deci, este derivabilă în orice sistem complet de logică propozițională.

Dar are cineva încredere în teorema acoperirii datorită acestei demonstrații? Mă îndoiesc. Și, în mod sigur, nimeni nu ajunge mai întâi să cunoască rezultatul, derivându-l realmente într-un sistem de logică propozițională. Demonstrația succintă – noncombinatorială – este cea care ne face să avem încredere în teorema acoperirii.



Doar pentru cazul în care exemplul cu tabla de șah obișnuită nu este convingător, să observăm cât de incomodă devine explozia combinatorială dacă mărim suprafața tablei. Totuși, demonstrația succintă funcționează la fel de bine pentru orice tablă  $n \times n$  (cu două colțuri opuse înlăturate), indiferent de mărimea lui  $n$ .

Acest exemplu ne trimite înapoi la latura platonistă a dihotomiei lui Bencerraf. Cunoașterea noastră a teoremei acoperirii nu se bazează pe o derivare. Abordarea combinatorială nu numai că eșuează în a face dreptate *adevărului*, dar nu-și atinge nici scopurile de *cunoaștere*. Într-o anumită măsură, cunoașterea noastră se bazează pe accesul intelectual la un gen de model. Este vorba despre un gen de relaționare care există în mod obiectiv, independent de noi și pe care, într-un fel sau altul, o putem sesiza. Roger Penrose exprimă aceasta cu strălucire: „... ori de câte ori mintea percepe o idee matematică, ia contact cu lumea lui Platon, a conceptelor matematice... Când cineva „sesizează“ un adevăr matematic, conștiința acestuia pătrunde în această lume de idei și ia contact direct cu ea...”<sup>4</sup>.

Platonismul conține mai multe ingrediente. (1) Obiecte abstracte existând în afara spațiului și timpului. (2) Felul în care sunt aceste obiecte este ceea ce face enunțurile noastre matematice adevărate sau false. (3) Mintea poate sesiza sau intui (unele dintre) aceste obiecte. (4) Cunoașterea matematică este a priori, în sensul că este independentă de simțurile fizice; dar nu este nevoie să fie infailibilă.

Prezentarea schițată este similară cu cea a lui Gödel, cel mai important platonist al zilelor noastre, care șcria:

... în ciuda îndepărtării lor de experiența sensibilă, noi avem ceva asemănător unei percepții a obiectelor teoriei mulțimilor, după cum reiese din faptul că axiomele ni se impun ele însele ca fiind adevărate. Nu văd nici un motiv pentru care ar trebui să avem mai puțină încredere în acest gen de percepție, adică în intuiția matematică, decât în percepția sensibilă<sup>5</sup>.

... presupunerea existenței unor asemenea obiecte este tot atât de îndreptățită ca și presupunerea existenței corpurilor fizice și avem tot atâtea motive să credem în existența acestora. Ele sunt necesare pentru obținerea unui sistem de matematică satisfăcător, în același sens în care corpurile fizice sunt necesare pentru o teorie satisfăcătoare a percepțiilor noastre sensibile...<sup>6</sup>.

Desigur, platonismul pe care l-am schițat se confruntă cu aceleași obiecții care i-au fost făcute lui Gödel; anume, cum este posibilă o asemenea „sesizare” sau „percepție”, atâta timp cât obiectele abstracte, fiind în afara spațiului și timpului, nu pot interacționa cauzal cu noi? Respingerea acestei obiecții este echivalentă, în sens larg, cu demonstrarea faptului că teoria cauzală a cunoașterii este greșită. Întrucât am arătat aceasta în altă parte<sup>7</sup>, voi trece la subiectul meu principal.

## EXPERIMENTELE MENTALE

Experimentele mentale se desfășoară în laboratorul minții. Dincolo de acest crâmpei de metaforă este greu de spus exact ce sunt ele. Le recunoaștem atunci când le vedem: ele sunt reprezentabile; ele presupun manipulări mentale; ele nu sunt numai consecințe ale unui calcul teoretic; ele sunt adesea (dar nu întotdeauna) imposibil de executat ca experimente reale, sau pentru că ne lipsește tehnologia relevantă, sau pentru că ele sunt pur și simplu imposibile în principiu.

Să examinăm acum cel mai frumos exemplu de experiment mental imaginat vreodată: admirabilul argument al lui Galilei din *Discorsi*, care arată că toate corpurile, indiferent de greutatea lor, cad cu aceeași viteză<sup>8</sup>. Argumentul începe prin menționarea concepției lui Aristotel, conform căreia corpurile mai grele cad mai repede decât cele mai ușoare ( $H > L$ ). Ni se cere apoi să ne

imaginăm că un proiectil greu de tun este atașat de un proiectil ușor de muschetă. Ce s-ar întâmpla dacă ele ar fi lansate împreună?



Fig.3

Raționarea în manieră aristotelică duce la o concluzie absurdă. Mai întâi, proiectilul ușor îl va încetini pe cel greu (acționând ca un fel de frână), astfel încât viteza sistemului combinat va fi mai mică decât viteza proiectilului greu, atunci când acesta cade singur ( $H > H+L$ ). Pe de altă parte, sistemul combinat este mai greu decât proiectilul de tun, așa că ar trebui să cadă mai repede ( $H+L > H$ ). Obținem acum consecința absurdă conform căreia proiectilul greu cade deopotrivă mai repede și mai încet decât sistemul combinat care este mai greu. Astfel, teoria aristotelică despre corpurile în cădere este distrusă.

Dar întrebarea rămâne: „Care cade mai repede”? Răspunsul corect este acum limpede ca lumina zilei. Paradoxul este rezolvat egalizând vitezele: toate cad cu aceeași viteză ( $H = L = H + L$ ).

Nu toate experimentele mentale funcționează în acest mod. Unele vizează numai distrugerea. Numesc experimentele mentale din această categorie, *distructive*. În mod tipic, experimentul mental distructiv este un gen de *reductio ad absurdum* al unei teorii preexistente.

Tânărul Einstein se imagina pe el însuși urmărind o rază de lumină și se întreba cum ar arăta aceasta dacă el ar putea alerga atât de repede. O rază de lumină este un câmp electromagnetic oscilant. Câmpul magnetic există din cauza unui câmp electric variabil, iar câmpul electric există din cauza unui câmp magnetic variabil – un fel de efect săritură-de-broască, în care broaștele există numai atâta timp cât sar. Dar dacă Einstein ar fi alergat alături de o astfel de undă, aceasta ar fi apărut ca fiind staționară, caz în care unda nici nu ar putea exista. Acest experiment mental este un splendid *reductio ad absurdum* al teoriei combinate a mecanicii clasice și electrodinamicii lui Maxwell.

Pisica lui Schrödinger este încă un exemplu de experiment mental pur distructiv. În acest caz ținta lui Schrödinger era interpretarea de la Copenhaga a mecanicii cuantice. Desigur, acest experiment mental nu arată că interpretarea de la Copenhaga este logic inconsistentă, așa cum procedează multe experimente mentale în raport cu țintele lor, ci mai curând că această interpretare violează flagrant reductibilul bun simț.

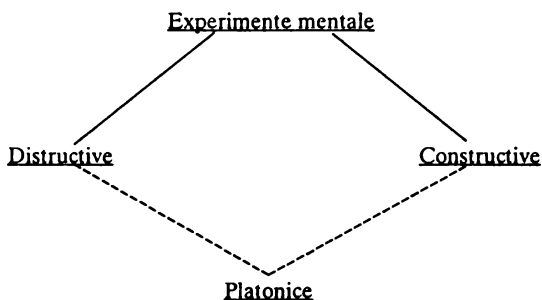
Multe experimente mentale fac parte dintr-o categorie complet diferită. În loc să joace un rol de respingere, ele furnizează dovezi în sprijinul unei teorii. Numesc experimentele mentale din această categorie, *constructive*<sup>9</sup>.

Newton imaginează o găleată cu apă într-un spațiu în alte privințe gol. În primul moment apa este nemișcată față de găleată, iar suprafața sa este plană; în al doilea moment există o mișcare relativă a apei și găleții; în al treilea moment apa și găleata sunt din nou nemișcate una față de alta, dar de data aceasta suprafața apei este concavă. Care este diferența dintre momentul unu și trei? Newton oferă următoarea explicație: în cel de-al treilea caz, dar nu și în primul, apa și găleata sunt în mișcare față de spațiul absolut. Prin urmare, spațiul absolut există.

Maxwell a fost campionul teoriei statistic-cinetice a căldurii, dar a trebuit să se confrunte cu problema de a nu putea să obțină a doua lege a termodinamicii într-o formă riguroasă, ci numai într-o versiune probabilistică. El nu putea să conchidă decât că este foarte puțin probabil, de exemplu, ca de la sine căldura să poată trece de la un corp rece la unul cald. Pentru unii, orice posibilitate ca așa ceva să se întâmple mirosează a absurditate. Experimentul mental cu demonul lui Maxwell a arătat că este într-adevăr posibilă trecerea căldurii de la un corp rece la unul cald. Demonul controlează o ușă care desparte două camere alăturate, una caldă, cealaltă rece. Viteza medie a moleculelor este diferită în cele două camere; câteva molecule rapide se află, de exemplu, în camera rece. Demonul le lasă pe acestea să pătrundă în camera caldă, în timp ce lasă moleculele lente să treacă din camera caldă în cea rece. În acest fel căldura a trecut din camera rece în cea caldă, deoarece viteza medie a moleculelor a crescut în camera caldă și a scăzut în cea rece.

Ca un exemplu final de experiment mental constructiv, să considerăm ascensorul lui Einstein. Observatorul este complet închis într-un ascensor și nu poate spune dacă ascensorul este nemișcat într-un câmp gravitațional sau se mișcă accelerat, în sus, într-un spațiu liber de forțe. Principiul echivalenței\* spune că nu există nici o diferență între aceste două stări. Să presupunem că ascensorul se află de fapt în mișcare accelerată și că o rază de lumină pătrunde printr-o mică gaură practică într-unul din pereți. Razei de lumină îi va trebui un timp finit pentru a traversa ascensorul, în timp ce acesta se mișcă în sus. Astfel, pentru observator, raza de lumină va apărea ca fiind curbată în jos. În consecință, conform principiului echivalenței, raza de lumină s-ar curba în același fel într-un câmp gravitațional.

Fizica este plină de experimente mentale care sunt sau distructive, sau constructive. Dar, oricât ar fi acestea de minunate, există o categorie mai deosebită de exemple. Numesc experimentele mentale din această ultimă categorie, *platonice*. Aceste experimente mentale sunt simultan constructive și distructive. Ele distrug vechiul și creează noul dintr-o singură suflare.



Există doar un mic număr de exemple din acest gen deosebit de experiment mental. Consider că argumentul lui Einstein, Podolsky și Rosen este un candidat posibil. Acest experiment mental a fost atât un *reductio ad absurdum* al concepției de la Copenhaga, cât și un argument pozitiv pentru teoriile cu parametri ascunși. (În

---

\* Dintre un câmp gravitațional și un corp inerțial (n. trad.).

acest punct aş sublinia caracterul failibil al experimentelor mentale. Faptul că teoriile cu parametri ascunşi sunt acum cunoscute ca fiind lipsite de speranţă nu afectează analiza modului în care funcţionează experimentele mentale). Totuşi, un alt posibil candidat este argumentul lui Leibniz pentru *vis viva*. În acelaşi timp, acest argument distruge concepţia carteziană despre ceea ce se conservă şi stabileşte o nouă perspectivă. Dar cel mai important exemplu dintre toate este admirabilul experiment mental al lui Galilei despre căderea liberă, pe care l-am schiţat mai sus. Pentru exemplificare, mă voi opri la acesta. Deşi raţionamentul în experimentul mental al lui Galilei este failibil, el nu este un fragment de cunoaştere a posteriori, ipotetică, empirică standard. Mai curând suntem îndreptăţiţi să numim aceasta un caz de cunoştinţă a priori. Iată de ce: În primul rând, *nu am avut nici o informaţie empirică*. Presupun că acesta este aproape un adevăr prin definiţie; fiind un experiment *mental*, se exclud noi intrări empirice. Nu este vorba de faptul că experimentul mental nu înglobează deloc informaţii empirice. Accentul este pus pe *noi* intrări senzoriale; aceasta este ceea ce lipseşte în experimentul mental. Ceea ce încercăm să explicăm este *tranziţia* de la vechea teorie la cea nouă şi aceasta nu este uşor de explicat în termenii intrării empirice, decât dacă există o nouă intrare empirică.

În al doilea rând, *noua teorie a lui Galilei nu este dedusă logic din informaţiile vechi. Şi nici un gen de adevăr logic nu este astfel dedus*. O a doua cale de a face noi descoperiri – o cale care nu-l deranjează pe empirist – este de a le deduce din informaţii vechi. Poate că experimentul mental al lui Galilei este într-adevăr un argument. Este acesta plauzibil? Cred că nu. Premisele unui astfel de argument puteau include toate informaţiile care erau admise în teoria lui Aristotel. Din acestea, Galilei deriva o contradicţie. (Până aici, totul este în regulă; avem un argument direct pentru acest punct). Dar putem noi să derivăm teoria lui Galilei, conform căreia toate corpurile cad cu aceeaşi viteză, din aceste aceleaşi premise? Ei bine, într-un sens da, deoarece putem deriva orice dintr-o contradicţie; dar aceasta pare greu de acceptat<sup>10</sup>. Mai mult, orice am putea deriva din aceste premise este imediat discutabil, deoarece pe baza contradicţiei considerăm acum cu îndreptăţire că încrederea noastră în premise este subminată.

Ar putea fi adevărată teoria lui Galilei numai în virtutea logicii? Pentru a vedea că teoria după care toate corpurile cad cu aceeași viteză nu este un adevăr logic, este suficient să remarcăm că, așa cum s-a pretins recent<sup>11</sup>, corpurile ar putea să cadă cu viteze diferite în funcție de culoarea lor sau de compoziția lor chimică. Aceste considerații subminează perspectiva argumentativă a experimentului mental<sup>12</sup>.

În al treilea rând, tranziția de la teoria lui Aristotel, la teoria lui Galilei nu este doar un caz de efectuare a celor mai simple și generale ajustări asupra teoriei vechi. Poate fi foarte bine cazul ca tranziția să fi fost cea mai simplă, dar nu acesta a fost motivul pentru a o face. (Mă îndoiesc că simplitatea sau alte considerații estetice joacă vreodată un rol util în știință, dar, de dragul argumentului, să admitem că pot juca un asemenea rol). Să presupunem că gradul de încredere rațională în teoria lui Aristotel despre căderea corpurilor este  $r$ , unde  $0 < r < 1^*$ . După ce a fost efectuat experimentul mental și a fost adoptată noua teorie, gradul de încredere rațională în teoria lui Galilei este  $r'$ , unde  $0 < r < r' < 1^{**}$ . Cu alte cuvinte, emit pretenția istorică după care gradul de încredere rațională în teoria lui Galilei exact după experimentul mental a fost *mai mare* decât gradul de încredere rațională în teoria lui Aristotel exact înainte de acest experiment (Remarcați timpii de evaluare. Evident, gradul de încredere rațională în teoria lui Aristotel *după* ce a fost descoperită contradicția se apropie de zero). Apelurile la noțiunea de revizuire minimă a încrederii n-ar duce nici măcar la un început de explicație a acestui fapt. Noi nu avem doar o nouă teorie – noi avem o teorie mai bună.

Pe lângă acestea, sunt și alte motive care amintesc de exemplul dat de cunoaștere<sup>13</sup> a priori a naturii; dar probabil cel mai interesant și mai speculativ are de-a face cu posibilele sale conexiuni, cu abordarea realistă a legilor naturii, propusă recent de Armstrong, Dretske și Tooley<sup>14</sup>.

Noua concepție pornește de la o insatisfacție legată de teoriile empiriste și nominaliste ale regularității. Hume și urmașii

---

\* Corect:  $0 \leq r \leq 1$  (n. trad.).

\*\* La fel  $0 \leq r < r' \leq 1$  (n. trad.).

săi moderni susțin că o lege a naturii este o regularitate. „Este o lege că toți corbii sunt negri“ este analizată ca fiind de forma  $(\forall x)(Cx \supset Nx)$ . Desigur, se recunoaște imediat că lucrurile nu stau chiar așa, deoarece există o mulțime de regularități care, cu siguranță, nu sunt legi ale naturii. „Toate persoanele din această cameră au monede de argint în buzunar“ corespunde modelului, dar cu greu am accepta că este o lege. Teoreticienii regularității, precum Ayer, Braithwaite și mulți alții, adaugă o a doua condiție pentru a marca diferența. Pentru a fi o lege, un enunț trebuie să fie mai întâi universal și apoi trebuie să joace un rol central în schema noastră conceptuală. Exemplul cu monedele-în-buzunar eșuează pentru acest din urmă motiv.

Cea mai evidentă obiecție se referă la subiectivitatea totală a concepției despre legi ca regularități. Pentru a fi o lege, din punctul de vedere al empiristului, cineva trebuie să o fi recunoscut ca atare; în consecință nu au existat legi înainte de a fi existat oameni. Persoane diferite pot avea diferite legi pur și simplu deoarece ele plasează diferite enunțuri universale (adevărate) în centrul țesăturii lor de opinii<sup>15</sup>.

Perspectiva realistă este în mai mare măsură în acord cu bunul simț; se spune că încrederea în aceste legi există independent de noi. Dar această perspectivă implică o angajare față de entități abstracte, adică față de universalii. Alături de corbi individuali și lucruri negre există universalii abstracte, proprietatea a-fi-corb (*ravenhood*),  $C$ , și proprietatea a-fi-negru (*blackness*),  $N$ , și acestea se află într-o relație de necesitate naturală,  $\mathcal{N}(C,N)$ . Această relație între universalii implică enunțul cuantificat universal care exprimă regularitatea, dar nu este implicat de acesta. Avem:

$$\mathcal{N}(C, N) \rightarrow (\forall x)(Cx \supset Nx) \text{ și totuși, } (\forall x)(Cx \supset Nx) \nrightarrow \mathcal{N}(C,N)$$

După Armstrong, Dretske și Tooley, această concepție despre legile naturii este metafizică pură. Ea dă socoteală despre natura legilor și explică regularitățile obținute. Dar modul în care învățăm noi despre legi este același ca și cel al empiriștilor înverdați: privim la instanțele individuale ale regularității; vedem corbii, niciodată proprietatea a-fi-corb.



Doresc să adaug un aspect epistemologic la această abordare metafizică a legilor. Sugestia mea este pur și simplu aceasta: în unele experimente mentale *sesizăm legile relevante* – nu regulăritățile, ci universalitățile însele. Dat fiind adevărul platonismului în matematică (adică, putem sesiza unele entități abstracte) și dat fiind că Armstrong ș.a. au ajuns la interpretarea corectă a legilor naturii (adică, legile sunt entități abstracte), n-ar trebui să ne așteptăm să întrezărim efectiv aceste legi? De fapt, nu ar fi puțin surprinzător dacă noi n-am fi apți să percepem, într-un fel sau altul, cel puțin unele din aceste legi?

În concluzie, să punem succint în contrast această formă de platonism despre fizică cu forme mai vechi de raționalism. Spre deosebire de Platon, Descartes, Leibniz etc., cunoașterea a priori în concepția mea nu este nici sigură, nici înnăscută. Ea nu a fost pusă acolo de Dumnezeu; nu este reamintire și nici nu este infailibilă. Dar, la fel ca raționaliștii tradiționali, susțin că tărâmul abstract este perfect real și că putem cunoaște câte ceva despre el. De asemenea, spre deosebire de raționaliștii tradiționali, concepția mea, conform căreia uneori dobândim cunoaștere a priori despre lumea fizică, nu este ea însăși a priori. Descartes, în opoziție, oferea un argument a priori pentru concepția sa după care cunoașterea noastră este a priori. (Era clar și distinct pentru el că ideile noastre clare și distincte trebuie să fie adevărate). Platonismul pe care îl apăr este o ipoteză – declar doar că această ipoteză este cea mai bună explicație a remarcabilului fenomen al experimentelor mentale. Acesta este un stil de argumentare pe care cei mai mulți filosofi contemporani îl găsesc acceptabil. Cu toate acestea, mi-am dat seama că ipoteza, în sine, este departe de a fi socotită astfel.

## NOTE

<sup>1</sup> Vezi, *The Laboratory of Mind*, Routledge 1991.

<sup>2</sup> Albert Einstein, „Geometry and Experience, retipărit în *Ideas and Opinions*, New York: Benanza Books, 1954, 275

<sup>3</sup> Paul Benacerraf, „Mathematical Truth“ retipărit în Benacerraf și Putnam (eds) *The Philosophy of Mathematics*, 2nd ed., Cambridge, 1983.

<sup>4</sup> Roger Penrose, *The Emperor's New Mind*, Oxford, 1989, p. 428.

<sup>5</sup> Kurt Gödel, „What is Cantor's Continuum Problem?“, retipărit în Benacerraf și Putnam (eds) *Philosophy of Mathematics*, Cambridge, 1983, p. 484.

<sup>6</sup> Kurt Gödel, „Russell's Mathematical Logic“, retipărit în Benacerraf și Putnam (eds) *Philosophy of Mathematics*, Cambridge, 1983, 456.

<sup>7</sup> Vezi J.R. Brown, „ $\pi$  in the Sky“, în A. Irvine (ed) *Physicalism in Mathematics*, Kluwer, 1989, și *The Laboratory of the Mind: Thought Experiments in the Natural Sciences*.

<sup>8</sup> Galileo Galilei, *Discourse on Two New Sciences*, University of Wisconsin Press, 1974, 66f (Vezi *Dialoguri asupra științelor noi*, București, Editura Academiei, 1961).

<sup>9</sup> Sunt posibile multe variații înăuntrul categoriei experimentelor mentale constructive, dar nu le voi urmări aici. O taxonomie completă este dată în *The Laboratory of the Mind: Thought Experiments in the Natural Sciences*.

<sup>10</sup> Altfel aş putea realiza experimentul acum și aş deriva „Luna este făcută din brânză verde“.

<sup>11</sup> De exemplu de către Fischbach ș.a., „Reanalysis of the Eötvös Experiment“, *Physical Review Letters* (Ian.) 1986. Bănuiesc că motivele pentru care experimentul mental al lui Galilei funcționează pentru ușor/greu, dar nu și pentru culori este că primele sunt aditive sau extensive, în timp ce ultimele nu (adică, prin combinarea a două obiecte roșii nu se obține un obiect de două ori mai roșu).

<sup>12</sup> Am fost prea grăbit și poate chiar incorect în respingerea perspectivei argumentative a experimentului mental. John Norton prezintă perspectiva argumentativă cu mai multă plauzibilitate. Examinez mai atent concepțiile sale în: *The Laboratory of the Mind: Thought Experiments in the Natural Sciences*.

<sup>13</sup> Termenul „Cunoaștere“ poate fi prea tare, întrucât implică adevărul, „încredere rațională“ ar putea fi mai bun deoarece, în concepția mea, ceea ce este a priori poate fi fals.

<sup>14</sup> Vezi David Armstrong, *What is a Law of Nature?*, Cambridge, 1983; Fred Dretske, „Laws of Nature“, *Philosophy of Science*, 1977; Michael Tooley, „The Nature of Laws“, în *Canadian Journal of Philosophy*, 1977. Bas van Fraasson, *Laws and Symmetry*, Oxford, 1989, conține multe observații critice importante la adresa abordării realiste a legilor, însă nu am să încerc să-i apăr pe Armstrong ș.a. aici.

<sup>15</sup> Pentru observații critice detaliate cu privire la concepțiile empiriștilor despre legi, vezi primele capitole din Armstrong, *What is a Law of Nature?* sau Tooley, *Causation: A Realist Approach*, Oxford, 1988.

Traducere Dumitru GHEORGHIU

# EXPERIMENTELE MENTALE ȘI EPISTEMOLOGIA LEGILOR\*

Roy A. SORENSEN

(Universitatea New York)

Scopul acestui articol este de a prezenta felul în care experimentele mentale ne ajută să învățăm despre legile naturii. După expunerea unor exemple de clarificare nomică, în prima secțiune, examinez explicațiile date cunoașterii noastre modale și optez pentru o abordare de tip evoluționist. Rezultatul fundamental constă în aceea că legile naturii ne conduc la dezvoltarea unor intuiții brute și rudimentare, despre posibilitatea fizică, exploatare ulterioară de către cei care fac experimente mentale pentru a dezvălui câteva din legile care sunt într-adevăr responsabile pentru aceste intuiții. Vesta cea bună este că selecția naturală asigură un anumit grad de plauzibilitate pentru intuiții. Vestea cea rea este că abordarea evoluționistă pare să limiteze domeniul experimentelor mentale plauzibile la contexte extrem de practice și concrete. În secțiunea a patra expun argumente în favoarea ideii că aceste limite nu sunt atât de severe pe cât sugerează interpretarea pesimistă a selecției naturale. Cu toate acestea, susțin, de asemenea, ideea că biologia poate asigura predicții și diagnoze utile privind eșecurile experimentelor mentale.

## I. EXPERIMENTELE MENTALE ȘI LEGILE

A. Experimentele mentale care nu vizează legi: Scenariile ipotetice nu trebuie să vizeze neapărat legi, pentru ca noi să învățăm

\**Thought Experiments and the Epistemology of Laws*, conferință ținută la Universitatea din Dubrovnik 1990 și publicată în *Canadian Journal of Philosophy* nr. 1, vol. 22, 1992.

din acestea despre legile naturii. Valoarea pedagogică a celor mai multe exerciții de manual provine din capacitatea lor de a permite o aplicație a legilor discutate în capitolul respectiv. Un experiment mental didactic îl pune pe student să-și imagineze că se află într-un tren, pe o șină fără frecare. Ce se întâmplă atunci când el începe să meargă pe culoar? Experimentul mental sugerează o aplicație a legii a treia a lui Newton, conform căreia pentru orice acțiune există o reacțiune egală și de semn contrar. Apoi, se presupune că trenul se află într-un tunel rectiliniu, de lungime infinită, astfel construit încât păstrează un vid perfect. Ce se întâmplă dacă studentul se oprește? Această întrebare de completare îl trimite pe student la o aplicație a primei legi a lui Newton, conform căreia un obiect va continua să se miște în linie dreaptă până va fi oprit de un obstacol. Oricum, legile lui Newton nu reprezintă *subiectul* acestor experimente mentale.

Alte experimente mentale vizează *legitatea*, mai curând decât o anumită lege. De exemplu, Michael Tooley arată că nu este nevoie ca legile să aibă un domeniu universal de aplicabilitate, imaginând o livadă ciudată<sup>1</sup>. Toate fructele din livadă sunt mere. Atunci când încerci să aduci o portocală în această livadă, ea se preschimbă într-un elefant. Bananele care trec hotarul livezii se preschimbă în mere. Plantează acolo un cireș și vei obține mai multe mere. Dată fiind această tendință puternică, se pare că există o lege conform căreia toate fructele din această livadă sunt mere – chiar dacă alte livezi similare nu se comportă în același fel.

B. Experimentele mentale nomice: Să ne întoarcem la experimentele mentale care vizează anumite legi. Aceste „experimente mentale nomice” clarifică expunerea legilor pe diferite căi.

1. Experimente mentale care au doar rolul de clarificare a legilor: un experiment mental nomic nu are neapărat rolul de a ataca sau de apăra o lege. Unele experimente mentale încearcă să stabilească o conexiune logică între două legi. Altele încearcă să caracterizeze legea în termeni metodologici; ele ne spun dacă legea este falsificabilă, redundantă, sau mai ușor de urmat. Alte experimente mentale încearcă să stabilească corelații psihologice. De exemplu, un cunoscut procedeu de clarificare este experimentul crucial ipotetic. Aristotel spunea că lucrurile grele tind să se deplaseze spre centrul Universului. Întrucât el susținea, de aseme-

nea, că centrul Universului coincide cu centrul Pământului, cineva ar putea avea îndoieli cu privire la intenția lui Aristotel de a formula legea despre un punct în spațiu sau despre Pământ. În cartea a patra din *De Caelo*, Aristotel își clarifică poziția, presupunând că Pământul ar fi urnit de la locul său din centrul universului. El afirmă că în acest caz, pietrele ar părăsi Pământul, deplasându-se către centrul universului. Acest experiment mental nu intenționa să justifice o lege, ci doar o pretenție psihologică privitoare la intenția cu care a fost formulată legea.

De asemenea, experimentele cruciale ipotetice clarifică legi punând în contrast consecințele acestora cu cele ale enunțurilor de legi rivale. William Gilbert și-a formulat „filosofia magnetică” în calitate de alternativă la filosofia aristotelică, revenind la ideea lui Empedocle, după care greutatea era o atracție între cele ce se aseamănă. După Gilbert, corpurile caută să se reunească, astfel încât lucrurile terestre tind spre Pământ, cele selenare spre Lună, iar cele solare spre Soare. Kepler nu era de acord cu această restrângere a acțiunii gravitației la lucruri omogene. El respingea, de asemenea, ideea lui Aristotel, după care un punct matematic poate avea semnificație fizică. Credința lui Kepler era că gravitația era o atracție reciprocă între oricare două corpuri. Aceste trei principii ale gravitației au fost puse în contrast cu ajutorul unui scenariu imaginat de Marin Mersenne<sup>2</sup>. Să presupunem că un om aruncă de pe Lună spre Pământ o piatră terestră și o piatră selenară. Conform principiului lui Aristotel, ambele pietre se vor deplasa spre Pământ. După principiul lui Gilbert, piatra terestră se va deplasa spre Pământ, în timp ce piatra selenară va cădea înapoi pe Lună.

2. Experimentele mentale care justifică legi: Mulți filosofi vor accepta imediat că experimentele mentale joacă un rol explicativ în cercetările nomice. Oricum, interesul meu principal privește experimentele mentale care confirmă sau resping enunțuri de legi aproape în același mod ca și experimentele obișnuite. Cea mai bună manieră de a stabili existența unei teze care justifică aserțiuni ipotetice este apelul la specimene uimitoare.

1. *Cazuri pure* : Cel mai direct tip de experiment mental de susținere, demonstrează o lege. Experimentul mental apare în mod firesc în acest proces, deoarece legile ne cer să examinăm

posibilități ce nu pot fi exemplificate în realitate. Această detașare de realitate a legilor este reflectată în remarca lui Arthur Eddington: „Abia dacă mai este nevoie să adăugăm că reflecția asupra unui domeniu mai larg decât cel real ne conduce la o înțelegere mai bună a realului”<sup>3</sup>. Oricum, nevoia de investigații contrafactice nu este întotdeauna apreciată. În 1930, R. A. Fisher se plânga că „pentru un matematician, afirmația (lui Eddington – nota aut.) este aproape o banalitate. Pentru un biolog, vorbind despre propriul său domeniu, ar sugera o perspectivă extraordinară de largă. Nici un biolog practicant interesat în reproducerea sexuală nu va fi condus să realizeze consecințele detaliate ale experimentelor pe organisme având trei sau mai multe sexe; totuși, ce altceva ar putea să facă dacă dorește să înțeleagă de ce sexele sunt, de fapt, întotdeauna două?”<sup>4</sup>. În continuare, Fisher schițează problemele coordonării organismelor trisexuate și dezavantajele reproducerii asexuate. După ce cititorul combină aceste două linii de speculație, el este mai puțin înclinat să vadă predominarea organismelor bisexuate ca un hazard statistic. Fisher aplică aceeași strategie problemei prevalenței proporției de 1:1 a sexelor. Se arată că alte proporții sunt exploatabile de către organisme care pot produce mai mult în economia sexelor.

Experimentele mentale ale lui Fisher sunt *explicative* în sensul că ele sugerează temeuri pentru un fapt acceptat. Ele merg puțin mai departe decât explicația, prin aceea că aprofundează nivelul vizibil de necesitate a faptelor. Alte experimente mentale nomice sunt *argumentative*, prin aceea că ele furnizează temeuri pentru o concluzie controversată. Un exemplu este anticiparea de către Galilei a legii inerției. Galilei a efectuat un număr de experimente obișnuite pentru a verifica legea înălțimilor egale: atunci când o bilă se rostogolește pe un plan înclinat, ea continuă să se miște, rostogolindu-se înapoi până la înălțimea inițială. În realitate, bila nu ajunge la înălțimea inițială, dar această discrepanță este imediat explicată ca o perturbație datorată frecării și rezistenței aerului. Experimentul ideal începe atunci când idealizăm aceste condiții și imaginăm bila rostogolindu-se succesiv pe planuri înclinate din ce în ce mai lungi, până când, la limită, vom avea un plan orizontal infinit. Legea înălțimilor egale presupune că bila se va rostogoli la nesfârșit, deoarece ea nu va ajunge niciodată la

înălțimea sa inițială. Acest rezultat atrage după sine o generalizare: un obiect în mișcare continuă să se miște în linie dreaptă. Generalizarea este interesantă deoarece vine în conflict cu ideea aristotelică potrivit căreia un obiect în mișcare tinde să se oprească.

Când vorbesc despre „susținere“ aici, nu am în vedere, desigur, un argument decisiv. Am în vedere numai sensul curent folosit atunci când vorbim despre modul în care experimentele obișnuite susțin ipoteze. Să ne oprim puțin pentru un scurt comentariu lămuritor. Eu nu cer nici măcar ca aplicația practică să-l convingă pe cel care face experimentul mental. Galilei însuși avea rezerve cu privire la intuiția la care făcea apel. Geocentrismul său perseverent l-a condus la credința că obiectele tind să se miște spre centrul Pământului, astfel că bila ar trebui să se deplaseze pe o mare circumferință. Dar, întrucât porțiunile scurte ale unei mari circumferințe nu se pot distinge de o linie dreaptă, Galilei a fost dispus să accepte generalizarea ca aproximativ adevărată.

Experimentele mentale sunt folosite, de asemenea, pentru a respinge enunțuri de legi. De exemplu, există o lungă istorie a credinței în principiul psihologic conform căruia orice alegere se bazează pe o preferință. Măgarul lui Buridan a fost prezentat ca un contraexemplu. Măgarul înfometat nu are nici o preferință între două grămezi de fân. Fără îndoială, el face o alegere *arbitrară*. Experimentul mental al lui Galilei despre căderea liberă a obiectelor alăturate este îndreptat împotriva unei *conjuncții* de legi aristotelice. Prima lege spune că obiectele mai grele cad mai repede decât cele ușoare. A doua spune că obiectul care cade mai repede va fi încetinit dacă este alăturat unui obiect care cade mai încet. Presupunând că un obiect greu și unul ușor sunt alăturate în mijlocul traiectoriei lor descendente, Galilei a arătat că aceste legi nu pot fi ambele adevărate. Conform primei legi, obiectul compus va cădea mai repede, deoarece el cântărește mai mult decât componentul greu. Dar, conform celei de-a doua legi, componentul greu (și, prin urmare, cel care cade mai repede, după prima lege) va fi încetinit când este alăturat celui ușor. Deoarece componentul greu nu poate să cadă simultan mai repede și mai încet, una din cele două „legi“ trebuie respinsă.

Un alt tip de experiment mental de respingere are doar rolul de a întări o obiecție, respingând un răspuns. Să considerăm legile

dialectice ale lui Aristotel despre mișcare. Bunul simț sugerează că obiectele se mișcă doar atâta timp cât sunt împinse sau trase. Teoria lui Aristotel a dezvoltat această idee și astfel a întâmpinat dificultăți în explicarea mișcării proiectilelor. Săgețile în zbor par a fi contraexemple, deoarece ele continuă să se miște chiar după ce părăsesc arcul. O soluție curentă a problemei era de a susține că aerul năvălește de la vârful săgeții la coada sa, pentru a preveni formarea unui vacuum. Astfel, aerul care năvălește împinge săgeata din spate. Jean Buridan critica această tentativă de soluționare pentru a-l apăra pe Aristotel. Să presupunem că ascuțim coada boantă a săgeții. Aceasta ar micșora suprafața părții din spate a săgeții și ar scurta timpul său de zbor. Dar noi știm că ascuțirea nu va avea, în realitate, acest efect. Astfel, obiecția inițială a fost repusă în drepturi.

2. *Modificări* : un experiment mental poate fi un amestec de respingere și confirmare: se respinge enunțul de lege original, dar se susține o versiune strâns legată de aceasta. Fleeming Jenkins a criticat principiul lui Darwin al variabilității speciilor, concepând spectacolul unui porumbel imaginar care încearcă să-și facă un cap din ce în ce mai mic. Problema era de a admite că speciile sunt varia-bile, dar de a nega că spațiul de variație este indefinit.

Adesea, cei care preconizează o respingere ajung, în cele din urmă, doar la o modificare. Paradoxul reversibilității al lui Loschmidt ilustrează acest model de atenuare dialectică. Experimentul mental viza încercarea lui Ludwig Boltzmann de a elabora o interpretare dinamică a principiului după care entropia este întotdeauna crescătoare. Boltzmann dorea să prezinte modul în care fizica newtoniană conducea la consecința că universul va deveni din ce în ce mai dezordonat. Loschmidt a respins acest proiect, imaginând un sistem de particule<sup>5</sup>. Particulele se află în repaus pe fundul unui container, cu excepția uneia care se află în repaus a deasupra celorlalte. Să o lăsăm acum să cadă. Impactul său va fi cauza unei reacții în lanț, astfel încât particulele devin din ce în ce mai dezordonate, până când ajung la o stare de echilibru static. Până acum, sistemul de particule se comportă exact așa cum prezic adepții entropiei ireversibile. Dar, acum, să inversăm vitezele tuturor moleculelor. Acesta face ca întregul proces să se desfășoare în direcție opusă, de la dezordine la ordine. Loschmidt



intenționată ca experimentul mental să respingă principiul entropiei crescânde. Dar acest experiment l-a convins pe Boltzmann că principiul trebuie să fie doar transformat într-o lege statistică și derivat din premise non-mecanice<sup>6</sup>. Astfel, în loc să facă apel la fizica indiferentă față de timp a lui Newton, Boltzmann a apelat la un fapt probabilistic: sistemele tind către dezordine, deoarece numărul de stări moleculare care contează ca dezordonate întrece cu mult numărul de stări care contează ca ordonate. Boltzmann semnală o analogie cu jocul de Loto cu cinci numere. Fiecare număr alcătuit din cinci cifre are o șansă egală de a fi extras, dar extragerea unui număr cu o dispunere dezordonată a cifrelor componente este mai probabilă decât extragerea unui număr ordonat, cum ar fi 12345.

## II. EMPIRISM ȘI INTUIȚII MODALE

Exemplele precedente arată că experimentele mentale ne pot instrui despre legi pe diferite căi. Cea mai controversată și misterioasă din aceste căi este justificarea. Căci, cum se poate ca o metodă *a priori* să ne învețe ceva despre adevăruri contingente? Empiriștii pot admite că experimentele mentale ne edifică asupra legilor logice, sau poate chiar asupra legilor matematice. Dar, se pare că doar un raționalist poate să acorde celui care face experimente mentale vreo autoritate asupra domeniului legilor naturii. Astfel, empiristul se prezintă ca și cum ar fi forțat să nege succesul aparent al experimentelor mentale justificatoare. Poate că o mică parte din aceste experimente pot fi puse în adevărata lor lumină de floricele retorice sau substitute pseudo-științifice pentru experimente reale. Dar un scepticism universal în raport cu astfel de cazuri intuitiv atrăgătoare ar fi fost tot atât de greu de acceptat, ca și un scepticism universal în raport cu judecățile morale. Este mai bine, deci, să urmărim o cale diplomatică și să încercăm să salvăm aparențele. Deci, să ne întoarcem la mecanismele posibile ale experimentelor mentale justificatoare.

A. Rolul credințelor (*beliefs*) modale: multe experimente mentale au o formă care amintește de *reductio ad absurdum*. Să considerăm argumentul lui Stevinus, după care orice masă dată de apă se află în echilibru când este scufundată în apă. Să presupunem că principiul nu ar fi adevărat. Să presupunem, în particular, că o masă de apă *A* coboară, deoarece apa ce o înconjoară nu-i dă sprijin. Întrucât *A* coboară, o altă masă de apă *B* îi va lua locul. Dar, deoarece *B* este fizic echivalentă cu *A*, *B* trebuie, de asemenea, să coboare, și astfel să fie înlocuită cu *C*. Acest proces continuă la infinit. Dar aceasta este mișcare perpetuă! Astfel, principiul trebuie să fie adevărat, deoarece negația sa conduce la o absurditate.

Dar ce fel de absurditate? Nu este o contradicție, așa cum cere forma argumentului *reductio ad absurdum*. Imposibilitatea este fizică, nu logică. Tot, astfel, nu este nici o absurditate logică în a conchide că măgarul lui Burdian va muri de foame. Orice experiment mental nomic poate fi sistematic dejucat dacă se acceptă anumite absurdități contingente. Chiar și atunci când absurditatea majoră este o imposibilitate logică, rămâne maniera lui Duhem de a adopta o absurditate minoră. De exemplu, aristotelicianul poate spune că experimentul mental al lui Galilei despre căderea liberă a obiectelor alăturate arată doar că este imposibil ca obiecte de greutate diferite să fie alăturate în mijlocul traiectoriei lor descendente. Cu toate acestea, nici una dintre aceste eschivări nu este convingătoare. Căci noi suntem în mod îndreptățit încrezători în intuițiile noastre despre ceea ce este posibil.

B. Sursa cunoașterii modale: o parte a cunoașterii noastre modale provine din teorii. Dar cunoașterea modală nu poate proveni în întregime din teorii, deoarece teoriile însele au nevoie de cunoaștere modală pentru a debuta. Deci, care este originea cunoașterii noastre preteoretice despre posibilități?

1. Răspunsuri greșite: Cineva s-ar putea salva cu un răspuns care reduce necesitatea la analiticitate. Convenționaliștii spun că propozițiile analitice sunt acele propoziții al căror adevăr se poate stabili pe baza regulilor limbajului. Apoi, ei tratează necesitatea întocmai ca pe un mod metafizic eronat de a vorbi despre

analiticitate. Astfel, propunerea lor este aceasta: cunoașterea noastră despre posibilitate se bazează pe stăpânirea regulilor lingvistice. Convenționalismul este concretizat în mod acceptabil. Dar este incomplet. Convenționalismul tratează doar necesitatea logică. El nu poate explica, însă, cunoașterea necesităților psihologice, biologice și fizice.

Eșecul acestei poziții spartane poate stimula apetitul pentru o propunere filosofică generoasă: oamenii au un simț modal. Acest simț ne permite să percepem alte lumi posibile în același fel în care percepem lumea externă reală. Sau, probabil, cel de-al șaselea simț al nostru detectează universalii sau alte suporturi de informație modală semnificativă<sup>7</sup>. Această viziune împărtășește slăbiciunile teoriei simțului moral și al teoriei lui Gödel despre simțul matematic. Unde este organul simțului? Ipoteza unui simț modal nu satisface criteriul fiziologului în atribuirea de simțuri.

În cazul în care credințele noastre modale ar constitui realitatea modală, nu ar mai fi nevoie de un organ de simț. Astfel, un adept al concepției subiectiviste despre modalități ar putea sugera că *p* este posibil atâta timp cât este crezut posibil de către un gânditor ideal. Gânditorii obișnuiți ar avea cunoaștere modală în măsura în care aproximează acest ideal. De exemplu, dacă idealul este problema de a fi rațional și informat, încrederea în opiniile noastre modale ar crește în aceeași măsură în care am elimina iraționalitățile și ignoranța. Ca orice teorie subiectivistă, această abordare a cunoașterii noastre modale este criticabilă, putând fi acuzată de circularitate. Căci, excepție făcând cazul în care „crezul posibil” este analizat independent de „posibil”, definiția utilizează chiar termenul pe care urmărește să-l definească.

A patra eventualitate, mult mai greu de tratat, este cea de a considera credințele noastre modale ca fiind cauza posibilităților și imposibilităților. De exemplu, dacă Dumnezeu plămădește posibilul ca reacție la năzuințele noastre cognitive, plauzibilitatea speranțelor noastre ar fi garantată. Ar fi chiar suficient dacă Dumnezeu ar servi drept o cauză obișnuită, punându-ne pe noi în armonie cu împrejurările modale. De exemplu, William Whewell gândea că singura cale de a explica cunoașterea geometrică și succesul științific în general, era de a accepta că Dumnezeu l-a creat pe om după chipul și asemănarea sa: „De ce alcătuirea lumii

ca obiect corespunde alcătuirii minții noastre, ca minte? În primul rând, deoarece ele sunt opera aceluiași Creator; dar acesta nu este singurul motiv; mai departe, deoarece alcătuirea lumii poartă semnul Gândurilor Minții Divine, iar mintea omenească este, parțial, părtașă la Gândurile Minții Divine. Și astfel nu vedem numai că obiectele există în spațiu și în relații spațiale, dar simțim că ele trebuie să existe astfel, iar acele relații spațiale sunt adevăruri necesare<sup>8</sup>.

Această viziune pune probleme dincolo de trimiterea la existența lui Dumnezeu. Așa cum admitea Whewell, nu se explică astfel cunoașterea înăscută la animale. Cum se face că un pui proaspăt ieșit din ou ciugulește insecte cu atâta precizie? Cum de știe albina să facă celule hexagonale (maximizând astfel cantitatea de miere ce poate fi stocată cu o cantitate dată de ceară)? Mai mult, nu există nici un mecanism plauzibil care să modeleze posibilul, pentru a-l adapta structurilor mentale. Este imposibil de modificat posibilul.

Poate că Dumnezeu a ales calea cea mai ușoară și a inversat ordinea priorităților: în loc să adapteze universul la mintea umană, a adaptat, probabil, mintea umană universului. Întrucât credințele noastre sunt contingente, evităm problema modificării adevărilor necesare. Oricum, rămân alte probleme. De exemplu, dacă Dumnezeu a făcut mințile noastre, de ce sunt greșite atât de multe din credințele noastre modale? Teoria explică lesne succesele proiectului mental, dar nu și eșecurile.

2. Răspunsul lui Darwin: Abordarea corectă a intuițiilor noastre modale a debutat odată cu răspunsul lui Darwin la obiecția lui Whewell, conform căreia teoria evoluționistă nu poate explica cunoașterea noastră a adevărilor necesare<sup>9</sup>. Soluția fundamentală la problema cunoașterii înăscute este un sintetic *a priori* al cunoașterii moștenite: „Platon (...) spunea în Phaidon că *ideile necesare* își au sorginea în pre existența sufletului, nu sunt derivabile din experiență (citește maimuțe în loc de pre existență).<sup>10</sup> Altfel spus, mecanismul „generează-și-elimină” al selecției naturale garantează că se poate moșteni cunoașterea din experiența altor animale. Astfel, puiul poate ciuguli cu precizie, deoarece alte păsări înaintea lui au prosperat în mod diferit, în concordanță cu credințele lor perceptive. Situația este analogă cu acumularea de experiență într-o îndeletnicire manuală. Multe

inovații eșuează, dar succesele accidentale se sedimentează în practici standard. Oamenii par să aibă puțină cunoaștere înăscută, dar ei sunt din plin înzestrați cu acea cunoaștere socială care provine din puterea noastră de imitație. Mai interesantă este contribuția adusă de imaginația noastră. Această capacitate ne ridică deasupra procesului necruțător care guvernează cunoașterea înăscută. Căci imaginația ne permite să dirijăm în gândire o parte din procesul încercării și erorii, astfel încât „rațiunea, iar nu moartea să respingă încercările greșite”<sup>11</sup>. Această tehnică a variației de substituție se bazează pe precizia intuițiilor noastre *a priori* despre ceea ce contează ca eroare. Încercările mentale care reușesc pot fi distinse de cele care eșuează doar dacă avem, cel puțin marginal, credințe solide cu privire la ce s-ar întâmpla dacă încercarea ar fi realmente executată. Din fericire, ceea ce este în mod semnificativ „a priori”, este relativ; este vorba de ceea ce poate fi cunoscut înaintea unei încercări anume, iar nu înaintea tuturor experiențelor, oricare ar fi acestea. Așadar, cel care face experimente mentale își poate suplimenta rezervele insuficiente de cunoaștere înăscută cu lecția pe care a învățat-o din experiența proprie, ca și cu cea achiziționată din relatările altora.

Acum, se prea poate ca Darwin să fi enunțat tot ceea ce Whewell aștepta cu nerăbdare – și fără nici o nădejde – de la teologie. Oricum, Whewell era un fundamentalist însetat de certitudine. Așa cum Darwin a înțeles treptat, natura nu face decât să selecteze acele caracteristici care sunt mai bune decât cele însușite prin competiție, iar nu *cele mai bune* caracteristici. Așadar, sunt puține speranțe să existe o corespondență perfectă între reprezentarea despre lume a unui organism și lumea reală.

W. K. Clifford a fost unul dintre primii gânditori care au analizat consecințele eronate ale epistemologiei evoluționiste. El nota că nimic nu garanta deplina adecvare a ideilor noastre despre spațiu și timp: „Cum pot eu ști că unghiurile unui triunghi sunt egale exact cu două unghiuri drepte, în orice împrejurări posibile, nu numai în acele regiuni ale spațiului unde s-a aflat sistemul solar, ci și oriunde în altă parte? Întreaga experiență a strămoșilor mei, acumulată de o sută cincizeci de milioane de ani, nu este mai demnă de încredere pentru a susține *aceasta*, decât

propria mea experiență a ultimelor cinci minute... Acum, însăși doctrina evoluției nu-mi îngăduie să admit vreo sursă transcendentă a cunoașterii; așadar, sunt împins să conchid, în ceea ce privește orice enunț aparent universal, sau că el nu este realmente universal, ci este un enunț particular despre sistemul meu nervos, despre mecanismul meu de gândire, sau că eu nu știu dacă enunțul este adevărat”<sup>12</sup>. Îndoielile lui Clifford în legătură cu geometria au fost eliberatoare, prin aceea că ele i-au stimulat interesul cu privire la geometriile non-euclidiene. Oricum, impresia clară lăsată de considerațiile evoluționiste este una de îngrădire. Noi apărem legați de strămoșii noștri, condamnați a mărșălui în interiorul unui cerc nu mai larg decât este necesar pentru a învinge în competiție. Darwin însuși scria în 1880: „Nu pot avea multă încredere, sau chiar deloc, în așa numitele intuiții ale minții omenеști care s-a dezvoltat... dintr-o asemenea minte pe care o posedă animalele”<sup>13</sup>.

3. Dezvoltarea de către Mach a epistemologiei darwiniste: Deși Darwin a descris în linii generale fundamentele epistemologiei evoluționiste, el nu a formulat destule detalii pentru a îndreptăți (cel puțin în raport cu etaloanele sale) publicarea rezultatelor. Din fericire, lipsa accesului la corespondența și carnetele de însemnări ale lui Darwin nu a împiedicat alte figuri marcante ale secolului al XIX-lea să sesizeze semnificația epistemologică a teoriei evoluționiste. Într-adevăr, Ernst Mach a aplicat în mod explicit descoperirile revoluționare din biologie la analiza problemei experimentului mental<sup>14</sup>. După Mach, selecția naturală face ca mințile să imite modele naturale. Acest mimetism este constant, în mare măsură inconștient și involuntar. Acest proces produce o bogăție de dispoziții mentale instructive. Totuși, această „cunoaștere instinctuală” este întrucâtva nearticulată, supusă erorilor și mai dogmatică în privința a ceea ce nu se poate întâmpla, decât în privința a ceea ce se poate întâmpla; „Cunoașterea instinctuală” ne înarmează cu un criteriu al adevărului. Experimentele mentale forjează această accepțiune a absurdității.

Ca empirist riguros, Mach s-a grăbit să accentueze că inaccesibilitatea proceselor ce au loc în spatele configurației cunoașterii instinctuale, creează iluzii raționaliste. Căci experimentele mentale par să dezvăluie o perspectivă care exclude cunoașterea

sintetică *a priori*. Verdictele creează sentimentul involuntar de amplificare a obiectivității lor aparente. Experimentele mentale pot să apară ușor ca fiind mai imperative decât experimentele obișnuite. Căci cel care face experimente mentale are un control remarcabil asupra variabilelor neesențiale; el poate, pur și simplu, să nu țină cont de neplăceri cum ar fi rezistența aerului. Procedura este *a priori*, deci nu trebuie să ne facem griji cu privire la erorile de percepție. Mai mult, logica din spatele cunoașterii instinctuale este ascunsă și astfel este sustrasă criticii. Unul din avantajele abordării naturaliste a experimentelor mentale, pe care o întreprinde Mach, este acela că suntem avertizați despre aceste iluzii și ni se furnizează un ghid asupra experimentelor mentale în care putem avea încredere.

Teoria evoluționistă a lui Mach despre experimentul mental merită dezvoltări ulterioare. O cale evidentă de a îmbunătăți teoria constă în modernizarea biologiei. În primul rând trebuie să ne îndreptăm spre o teorie evoluționistă corectă. Alfred Russel Wallace gândea că putem explica dezvoltarea intelectuală umană numai postulând o „Inerție evoluționistă“, care ne-ar trimite dincolo de limitele contextelor practice<sup>15</sup>. Întrucât această forță a fost discreditată, trebuie să recurgem la o altă explicație a modului în care oamenii au putut să-și dezvolte capacitatea de a face calcule, de a demonstra teoreme și de a înțelege legile care guvernează universul. Va trebui să renunțăm chiar la forțe pe care Darwin însuși le-a utilizat, cum ar fi moștenirea dispozițiilor mentale<sup>16</sup>. O epistemologie evoluționistă adecvată trebuie să îmbogățească modelul Darwin-Mach cu detalii ce provin din biologia *contemporană*. O asemenea epistemologie trebuie să se limiteze la a accepta și desfășura forțe evoluționiste cu sensibilitate la ceea ce este biologic plauzibil. Deci, este încurajator să știm că epistemologia evoluționistă a fost „în aer“ în biologie. Mulți biologi au argumentat că mintea, ca orice altă parte a corpului, s-a format astfel încât să corespundă lumii. Konrad Lorentz scria: „Acest aparat nervos central nu prescrie legile naturii mai mult decât prescrie copita calului forma terenului. Ca și copita calului, acest aparat nervos central se împiedică de schimbări neprevăzute în ceea ce are de făcut. Dar, așa cum copita calului este adaptată la terenul stepei cu care se confruntă, tot așa, pentru a organiza imaginea despre lume, aparatul nervos

central este adaptat lumii reale cu care se confruntă omul. Ca orice alt organ, acest aparat a atins forma adecvată și conservată la nivelul speciei prin această copiere a realului de către real, de-a lungul multor epoci ale istoriei unei specii<sup>17</sup>. Oricum, adaptarea nu presupune unicitatea. Tot așa cum pot exista variații în privința copitelor, pot exista variații în privința minților, chiar dacă fiecare minte reflectă aceleași legi ale naturii. Donald Campbell accentuează aceasta printr-o analogie<sup>18</sup>. Legile hidrodinamicii plus avantajele locomoției au forțat peștii, balenele și morsele să ia aproximativ aceeași formă – o formă ce exprimă „cunoașterea” acestor legi. Dar, aceleași legi au modelat calmarii după o formă tot atât de „precisă”, dar radical diferită.

### III. BIOLOGIA DINCOLO DE IPOTETIC

Proiectul lui Darwin ne oferă, cu siguranță, un punct de plecare pentru abordarea experimentelor mentale. O teorie evoluționistă a intuițiilor modale evită dificultățile proiectului divin. În loc să se bazeze pe ideea că Dumnezeu modelează mintea, această teorie se bazează pe forțe de selecție deja acceptate de biologi. Aceste forțe sunt lipsite de capacitatea de anticipare și trebuie să lucreze cu limitările ingineresti ale materialului disponibil, astfel încât credințele modale rezultate au, cu siguranță, imperfecțiuni realiste. Totuși, ele trebuie să fie îndeajuns de plauzibile pentru a fi atât de utile încât să fie dezvoltate și conservate. În anumite împrejurări, o credință falsă are consecințe superioare în raport cu o credință adevărată. Dar, în general, adevărul este mai avantajos.

Credințele adevărate merită să fie reținute deoarece ne ajută să luăm decizii fecunde. Greșelile în legătură cu prădătorul și prada, împerecherea și rutele de migrație, hrana și adăpostul, conduc la alegeri în condițiile unei informații precare. Decizia de a vâna păsări săpând o capcană nu este o idee bună, deoarece omite posibilitatea ca pasărea să zboare din capcană. A te ascunde de vulturi este o precauție inutilă, întrucât vulturii nu mănâncă



decât cadavre. În general, adaptarea pregătește pentru posibilități autentice și nu cheltuiește energie pentru imposibilități. Atunci când adaptarea este necorespunzătoare, refacerea este rapidă, având loc prin intermediul unei concepții despre următoarea și cea mai probabilă eventualitate. Deci, capacitatea individului de a-și corecta și extinde imaginea despre lume va trebui să se difuzeze la nivelul posibilității de decizie a speciei. Raționamentul ipotetic nu este un turn de fildeș de lux. Este o necesitate impusă de legea junglei.

Unele raționamente ipotetice sunt pur extensionale, adică reprezentabile în termeni de condiționali materiali. Pentru a arăta că opiniile modale sunt utile, avem nevoie să demonstrăm utilitatea condiționalilor *optativi*. Căci numai acești „contrafactuali” ne constrâng să vorbim despre posibilități. Logicienii au un exemplu preferat care arată nevoia de a distinge între condiționalii indicativi și cei optativi:

(1) Dacă Oswald nu l-a împușcat pe Kennedy, atunci altcineva a făcut-o.

(2) Dacă Oswald nu l-ar fi împușcat pe Kennedy, atunci altcineva ar fi făcut-o.

Condiționalul indicativ (1) este evident adevărat, dar cel optativ (2) este controversat. Adevărul lui (2) ar presupune existența unei conspirații teroriste și ar avea o semnificație politică.

Criteriul exact al distincției dintre condiționalii indicativi și cei optativi este controversat. Cei care privesc condiționalii indicativi ca fiind semantic (dar nu și pragmatic) echivalenți cu cei materiali, consideră că acest contrast este considerabil. De exemplu, Frank Jackson și David Lewis sunt diversificaționiști. Ei susțin că acești condiționali indicativi sunt verifuncționali și se referă la lumea reală; condiționalii optativi nu sunt verifuncționali și se referă la lumi posibile. În plus, asertabilitatea unui condițional indicativ diferă de probabilitatea sa. Lewis considera că diferența dintre cele două tipuri de condiționali rezidă în esența problemei lui Newcomb: doi jucători citesc condiționalii care descriu matricea de plăți a jocului ca optativi, în timp ce un singur jucător îi interpretează ca indicativi. Unificaționiștii, cum sunt Brian Ellis și Robert Stalnaker, nu consideră condiționalii indicativi drept condiționali materiali și văd puține diferențe între

indicativi și optativi. Deși problema explicării utilității ipoteticului apare mai acut la diversificaționiști, ea apare într-o formă voalată și la unificaționiști. Căci, atunci când Ellis și Stalnaker accentuează asemănarea dintre indicativi și optativi lor aparent inutili, ei presupun problema utilității condiționalilor indicativi.

Deci ambele părți s-ar putea consola cu faptul că propoziția condițional-optativă (2) *contrastează* cu optativi inutili, precum: „Dacă Napoleon ar învia din morți, atunci testamentul său nu ar mai fi valabil“. Condiționalii optativi pot avea propria lor semnificație practică în aceeași măsură ca și propozițiile declarative. În plus, mulți optativi au o mai mare semnificație practică decât multe propoziții declarative. Comparați „Orbita planetei Neptun este eliptică“ cu următorii contrafactuali:

(3) Dacă nu ar fi fost foc la gura peșterii, acel urs ar fi intrat în peșteră.

(4) Dacă acei bizoni ar fi fost speriați, ei ar fi fugit în panică peste stânci.

(5) Leopardul ar fi furat hrana, chiar dacă ea s-ar fi aflat în copac. Contrafactualii sugerează precauții și prilejuri favorabile și le înfățișează pe cele nefavorabile. Prin urmare trebuie să respingem mitul filistin după care ipoteticul este mai puțin util decât relatarea faptelor necondiționate.

Ideea centrală a raționamentului optativ este reflectată chiar în vocabularul de bază al raționamentului primitiv. Căci acest vocabular abundă în termeni dispoziționali, cum sunt: „prieten“, „comestibil“, „prădător“, „adăpost“. A descrie o ciupercă otrăvitoare înseamnă a admite condiționalul optativ „Dacă ar fi să mănâncă cea ciupercă, m-aș îmbolnăvi“. Spre deosebire de condiționalul material corespunzător, optativul este fals dacă ciuperca este în realitate comestibilă, dar se întâmplă ca eu să nu o mănânc.

Curiozitatea atestă, de asemenea, importanța credințelor modale. Căci activitățile de explorare nu sunt creditate întotdeauna cu un câștig imediat, sau chiar cu un câștig realizabil într-un viitor apropiat. Așa cum nota Lorenz, problema este de a învăța ceva ce *poate* fi valoros. „Corbul care cercetează un obiect nu are dorința de a-l mânca; șoarecele care examinează toate ungherele și crăpăturile accesibile lui nu are dorința de a se ascunde; ambii vor să știe dacă obiectul respectiv *poate* fi mâncat sau utilizat ca ascunză-toare“<sup>19</sup>. Jocul poate fi, de asemenea, un vehicul al

cunoașterii modale. În mare măsură, jocul are calitatea unei repetiții și a unei încercări-limită. Tot așa cum un pilot de încercare este plătit să cerceteze „învelișul exterior“ al unui avion, animalele profită când află cât de repede pot să alerge și cât de puțin zgomot pot face.

Animalele se comportă, de asemenea, în moduri care îndreptățesc să li se atribuie idei abstracte. Darwin arăta că un câine care reperează un alt câine la distanță, îl sesizează pe acesta în mod abstract. Căci, atunci când câinele care se apropie este eventual recunoscut ca prieten, asistăm la o modificare de comportament. De asemenea, un semn al stăpânului poate să păcălească un câine, făcându-l să-și închipuie un anumit vânat. Când cercetările inițiale ale câinelui pe sol eşuează, el se uită în copaci după o veveriță imaginară, iar nu după o veveriță reală.<sup>20</sup>

Animalele obțin succese cognitive prin înzestrări naturale și instruire. Probabil, unele credințe, cum ar fi frica de întuneric, au o determinație genetică specifică. Dacă așa stau lucrurile, o parte a cunoașterii noastre modale poate fi înnăscută. Într-o abordare verosimilă, procesul generează-și-elimină al evoluției ar fi mecanismul de formare al credințelor plauzibile. Această cale de acces spre cunoaștere ar avea același statut ca și percepția, memoria și învățătura din relațiile altora. Controlul genetic asupra altor credințe poate fi mai slab, reprezentând doar o dispoziție de a forma o anumită clasă de credințe. Din perspectiva instruirii, părinții își educă și își învață copiii. Nu este nevoie de vreun plan, în spatele acestor lecții. Descendenții învață din așteptările celor mai experimentate animale și astfel absorb multă cunoaștere modală, fără vreo instrucție premeditată.

Prin urmare, teoria evoluționistă sugerează o explicație naturalistă a felului în care obținem cunoaștere modală prin intermediul experimentului mental. Nu a existat niciodată vreo îndoială că anumite categorii de minți pot avea credințe modale corecte. (Imaginați-vă, numai, un geniu modal apărând pe neașteptate). Problema era de a explica de ce apariția unei asemenea minți nu are nevoie de miracol. Abordarea evoluționistă rezolvă problema arătând cum natura ne pune, în mod constant, în acord cu ea însăși. Plauzibilitatea conferită credințelor noastre modale ne asigură că experimentele mentale care le exploatează vor moșteni această caracteristică.

Garanția nu este, oricum, universală. Căci există variabilitate în recompensele pentru credința în adevăruri modale și în pedepsele pentru credința în falsități modale. Jaloanele sunt mai înalte pentru obiecte de mărime mijlocie care se modifică cu rate de nivel mijlociu. Aici, nevoia de reușită poate explica popularitatea experimentelor mentale în mecanică. Diferențele de costuri ale erorilor ne sugerează, de asemenea, că experimentele mentale despre scenarii-limită sunt mai puțin plauzibile decât acelea despre obiecte familiare, la intensități familiare. În general, ne vom aștepta ca experimentele mentale să funcționeze cu succes numai în măsura în care prin ele se obține o consecință concretă dintr-o poziție teoretică.

Intuițiile modale au apărut ca un ajutor în luarea de decizii practice, deci ne vom aștepta ca ele să se îndrepte către acest scop. Astfel, ele se vor aplica imediat în contexte normative. Într-adevăr, vom găsi o mulțime de experimente mentale în etică, teoria deciziei și teoria argumentării juridice. Accentul practic conduce, de asemenea, la o diferență structurală. Nu există credințe adevărate accidentale.

1. O strategie deductivă care produce, în general, credințe adevărate, este o *strategie deductivă rațională*.

2. Selecția naturală produce strategii deductive raționale<sup>21</sup>. Dacă acest argument este corect, eșecul cognitiv se poate datora doar ghinionului sau bolii. De aici rezultă că studiul erorilor este greșit condus. Prin definiție, un eșec cognitiv este o eroare obișnuită de raționare. Dacă evoluția face din aceste erori evenimente întâmplătoare, atunci nu există eșecuri cognitive.

Stephen Stich a obiectat că prima premisă a raționamentului este falsă deoarece selecția naturală favorizează uneori strategiile deductive care produc în mare măsură credințe false. El respinge, de asemenea, și a doua premisă, pe temeiul că strategiile care produc o mare cantitate de adevăruri sunt uneori neraționale (deoarece ele pot fi de o abundență epistemică risipitoare). Întrucât sunt interesat în special de plauzibilitatea raționamentelor noastre modale mai curând decât de optimizarea lor, mă voi concentra asupra unor exemple de ilogicități naturale.

B. Căi pe care pot evolua eșecurile cognitive: Adevărul este doar un mijloc pentru atingerea țelului Mamei Natură. Ea maximizează succesele fertile, iar nu utilitățile epistemice. Deci există

spațiu de joc pentru eșecuri cognitive oriunde cererea de fertilitate se abate de la cererea derivată de credințe corecte.

1. Selectarea de noi reguli neplauzibile: Există câteva căi de a proceda greșit. Cea mai directă cale pentru selecția naturală este de a ne păcăli de la început cu o regulă inadecvată.

a. *Indisponibilitatea regulilor adecvate* : Există multe caracteristici utile pe care animalele nu le pot dezvolta. O alegere evidentă pentru o culoare de protecție este verdele, dar nici un mamifer nu are blană verde. Aceasta deoarece, probabil, caracteristica este indisponibilă, deci selecția naturală trebuie să aleagă între variante mai puțin eficace. Se poate ca în anumite cazuri meniul să fie într-adevăr limitat, forțând alegerea unei reguli care este doar cea mai bună dintr-o mulțime de reguli inadecvate. De exemplu, limbajul omenesc putea să aibă o forță limitată de expresie și astfel să nu fie apt să exprime cu precizie anumite relații de rudenie. Astfel, putem fi forțați să înregistrăm o corelație în termeni primitivi cum ar fi măsurarea vârstei copiilor cu ajutorul înălțimii lor.

b. *Inaccesibilitatea regulilor adecvate*: Chiar dacă o regulă plauzibilă este disponibilă, ea poate fi inaccesibilă. Regulile complicate pot fi dificil de codificat. Prin urmare, o regulă de deducție cu calități minore poate fi preferată unei reguli plauzibile mai costisitoare. Mama Natură va face cumpărături în același spirit ca și zgârciții care au nevoie de termometre, cadrane gradate și manometre.

c. *Compensări*: Uneori, progresele într-un anumit domeniu provoacă regrese în alte domenii. Reglarea irisului îmbogățește contrastul și constanța culorii, reducând efectul extrem de neplăcut al variației de iluminare. Din nefericire, mecanismul compensatoriu ne face vulnerabili la iluzii de contrast.

d. *A greși pentru a fi în siguranță*: Eroarea de a lua hrana otrăvitoare drept comestibilă este mult mai costisitoare decât eroarea de a lua hrana comestibilă drept otrăvitoare. Există mai multă pasiune selectivă împotriva erorilor anterioare. Și, într-adevăr, experimentatorii au găsit că șoarecii își formează rapid aversiune față de diferențele de gust ale hranei supuse unor radiații otrăvitoare<sup>22</sup>. Prin urmare, eroarea „generalizării pripite“ îi servește bine pe șoareci.

e. *Legături, cu caracteristici utile*: Pleiotropia este un fenomen obișnuit: o genă produce două caracteristici fenotipice dintre care numai una este utilă. Legătura poate să apară, de asemenea, la nivelul genotipului; o genă asociată cu o caracteristică utilă. În ambele cazuri, una dintre caracteristici este tolerată, chiar dacă ea nu ajută organismul. Într-adevăr, ea poate fi chiar letală. Astfel, o strategie deductivă care produce clase de credințe false poate fi tolerată de selecția naturală.

f. *Utilitatea epistemică extrinsecă*: Roderick Firth a trasat o distincție între valoarea epistemică intrinsecă și cea extrinsecă. Valoarea intrinsecă a unei credințe este determinată de valoarea sa de adevăr. Valoarea extrinsecă este determinată de valoarea sa epistemică. O intuiție poate conduce la achiziția unor credințe adevărate, având astfel mai mult valoare extrinsecă, decât intrinsecă. Același lucru poate fi spus despre o falsitate fertilă. Într-adevăr, de ce valoarea sa epistemică extrinsecă n-ar putea compensa lipsa de valoare intrinsecă? Probabil aceasta explică dispoziția noastră de a avea încredere în enunțuri de legi, în ciuda faptului că ele au fost odată respinse.

2. *Reguli degenerate și îmbătrânite*: Diminuarea costurilor nu este limitată la selectarea de noi reguli. Să considerăm „stimulii supranormali“ ai păsărilor. Fluierarul își indentifică ouăle prin *contrastul* înfățișat de petele cafenii pe un fond cafeniu deschis. Astfel, el preferă un ou artificial cu pete cafenii pe fond alb, deoarece contrastul este mai mare. Vânătorii de stridii pot fi păcăliți, fiind puși să clocească ouă de dimensiuni foarte mari, mai curând decât ouă de dimensiuni normale. În loc să răspundă stimulilor complicați ai unui pui flămând, anumite păsări produc provizii de hrană pentru a-și simplifica sarcina.

Există o presiune selectivă împotriva conservării regulilor „superplauzibile“. Mama Natură va fi întotdeauna în căutare de înlocuitori necostisitori. Atracția noastră pentru dulciuri poate fi un exemplu. Strămoșii noștri au fost astfel conduși să mănânce fructe coapte și hrănitoare și să evite fructele verzi și acre. Dar acum atracția pentru dulciuri ne conduce să preferăm zahărul, fructelor coapte.

3. *Extrapolarea unei strategii deductive preexistente*: Stich arată că există o specie de broaște care pot învăța să evite un miriapod periculos după numai o singură încercare. Oricum,

aceeași broască va mânca fasole până va deveni un sac viu cu fasole. O explicație este că regula sa de deducție este: „Lucrurile bune la gust sunt comestibile“. Această regulă funcționează bine în mediul său natural, dar nu și în laborator. Probabil că mitul după care cei mai mulți șerpi sunt periculoși se datorează unei reguli de deducție formată în cele mai îndepărtate timpuri, când, într-adevăr, cei mai mulți șerpi erau periculoși. Atunci când strămoșii noștri au migrat, ei au păstrat regula de deducție și în zone în care cei mai mulți șerpi erau inofensivi.

În alte cazuri, problema este mai curând timpul decât spațiul. Stich speculează că „plasticitatea euristică primitivă“ discutată de psihologi este o relictă din stadiul de vânători-culegători al strămoșilor noștri. În vremuri de demult, gândirea prin analogie pe care o asociem cu superstiția și pseudoștiința funcționa bine. Dar, întrucât condițiile s-au schimbat, acest stil de gândire este perimat.

4. Raționalitatea ca un obstacol pentru fertilitate: Adepții eugeniei se plâng că inteligența a devenit invers proporțională cu fertilitatea. Femeile inteligente evită familiile extinse, deoarece o viață dedicată creșterii copilului este nesatisfăcătoare din punct de vedere intelectual. Femeile slab înzestrate intelectual sau nu-și bat capul cu asta, sau nu sunt capabile să evite o astfel de viață. Ca urmare, populația devine din ce în ce mai stupidă. Adepții eugeniei mai spun că practici religioase precum celibatul pot tria inteligențele.

5. Forțe evoluționiste alternative: Există forțe evoluționiste dincolo de selecția naturală. Unele dintre aceste forțe nu conferă plauzibilitate credințelor. Într-o comunitate izolată, tendința genetică poate modifica strategiile deductive, astfel încât ele să nu mai fie eficace. Selecția sexuală poate conduce femeile să prefere masculii care au mulți mușchi și puțină minte.

#### **IV. PLEDOARIE PENTRU UN OPTIMISM MODERAT**

Selecția naturală poate produce indivizi lipsiți de rațiune. Să ne imaginăm apariția unui virus care omoară pe toți indivizii

dotati cu rațiune și numai pe aceștia. Sau să presupunem că oamenii devin paraziți ai altor organisme. Deoarece aceștia absorb hrana numai de la gazde, ei se vor simplifica în maniera teniilor. Astfel, viața lor mentală, devenind un epifenomen, va fi dominată de eroare înainte de a se preface în incoerență.

A. Răspunsuri pentru un pesimist: Optimismul epistemic extrem este respins de posibilitatea lipsei sistematice de rațiune. Oricum, moderații au suficient spațiu de manevră pentru a neutraliza obiecțiile pesimiștilor.

1. Contraexemplul toleranței: Faptul că evoluția nu presupune logic raționalitatea arată fără șovăire că evoluția nu poate explica succesele noastre intelectuale. Existența unei piețe libere nu atrage după sine practici de afaceri eficiente, dar aceasta nu le face pe cele două să nu aibă nici o legătură una cu alta. În definitiv, suntem liberi să adăugăm premise noi. Putem slăbi concluzia. Și ne putem mulțumi cu o conexiune inductivă, mai curând decât cu una deductivă. În particular, putem argumenta că evoluția mărește probabilitatea ca prezența calității de om rațional să fie plauzibilă pentru majoritatea populației. Prin urmare, pretenția ca evoluția să determine formarea unor credințe plauzibile trebuie luată în același mod ca și următoarele generalizări: practica determină îmbunătățirea performanțelor, cinstea conduce la reputația de a fi cinstit, celibatul implică lipsa de boli venerice. Nici unul din aceste enunțuri cauzale generice nu poate fi respins menționând împrejurări particulare în care conexiunea nu are loc. Există chiar situații în care, în mod *sistematic*, conexiunea nu are loc. De exemplu, uneori practica slăbește performanța, producând oboseală, pagube sau performanțe inutile.

2. Legătura dintre practică și teorie: Așa cum arăta Elliot Sober, raționalitatea umană va apărea ca un mister dacă se tratează caracteristicile practice și cele intelectuale ca reciproc exclusive. Epistemologul evoluționist va insista asupra faptului că există legături strânse între multe abilități practice și teoretice.

Sobers a produs chiar argumente pentru a gândi că există o conexiune *sistematică*. Strânsa întrepătrundere a strategiilor deductive cu împrejurările concrete ar cere un nivel ridicat de detalii pe care regulile ar trebui să le redea. Aceasta ar încălca memoria și ar face dificilă aplicarea regulii. Deci selecția naturală ar favoriza regulile simple pentru a ușura sarcina de a le urma. De aceea,



caracterul practic conduce selecția naturală la maximizarea capacității de soluționare teoretică. Celelalte exigențe teoretice – consistența, completitudinea și verificabilitatea – se îmbină cu caracterul practic. Nici nu e nevoie să fim atât de subtili – pentru a aprecia caracterul practic. Nici nu e nevoie să fim atât de subtili pentru a aprecia caracterul nepractic al imperfecțiunilor teoretice ca o circularitate.

3. Dinstincția dintre selecție *a* și selecție *pentru*: Mama Natură este zgârcită în sensul că selectează caracteristici ce ajută un organism numai pentru a învinge în competiție. Oricum, o selecție *a* unei caracteristici nu presupune selecția *pentru* o caracteristică. Deci, mama natură nu este zgârcită în raport cu acele caracteristici avantajoase *per se*. Astfel, „surplusul” de capacitate mentală nu este *automat* o anomalie.

4. Descrieri ale unor caracteristici problematice: Atribuirea de credințe este o problemă delicată. Stich descrie șoarecii iradiati ca și cum ei ar crede că bucățelele de pâine cu diferite arome sunt otrăvitoare. De ce să nu fim mai îngăduitori și să-i descriem ca și cum ar crede că acele bucățele *pot* fi otrăvitoare. Această credință rațională este puțin mai complicată, deoarece conține un element modal. Dar dacă suntem gata să admitem că șoarecele este destul de sofisticat pentru a înțelege termenul dispozițional „otrăvitor”, nu vom fi șocați dacă îi atribuim credințe modale de tipul „poate”.

5. Redundanță: Psihologilor percepției le place să caracterizeze imaginea drept un „sac cu iluzii”. Reușim să dăm un sens la ceea ce vedem utilizând un grup larg de reguli empirice acoperitoare. Fiecare regulă are slăbiciunile sale care pot fi izolate în manieră distructivă. Dar, atunci când aceste reguli acționează împreună, așa cum se întâmplă în mod normal, ele sunt plauzibile, deoarece sunt puține situațiile în care regulile dau greș simultan și în aceeași direcție. Lecția ce poate fi desprinsă de aici este că trebuie să evităm eroarea compoziției: de la implauzibilitatea părților nu putem conchide implauzibilitatea întregului.

6. Diferențele care îl despart pe om de animal: Pe lângă asemănările izbitoare dintre oameni și animale, trebuie să considerăm și diferențele. Multe organisme sunt specializate. Aceasta face ca ele să fie limitate cognitiv. Omul este un generalist între primat (la fel ca șoarecele între rozătoare și corvidul între păsările cântătoare). Generaliștii sunt făcuți să exploateze o largă varietate de condiții și astfel să aibă orizonturi relativ largi.

Indivizii umani sunt capabili de înnoire; tinerețea lor este prelungită și ei rețin permanent un mare număr de trășături din tinerețe. Deoarece tinerețea este perioada căutărilor și a progresului cognitiv, ființele umane surclasează intelectual celelalte animale. Oamenii au, de asemenea, avantajul limbajului, o durată relativ îndelungată de viață și libertate de acțiune. Astfel oamenii au, după cum se pare, mai puține puncte slabe și o capacitate mult mai mare de a compensa insuficiențele cognitive. Într-adevăr, diferențele dintre om și animal sunt mai curând de grad, decât de natură, dar adesea este vorba de diferența dintre a arunca un proiectil și a împușca cu un proiectil.

B. Depășirea limitelor de vânător-culegător: Am văzut deja cum unele limite aparent impuse de moștenirea noastră de vânători-culegători sunt iluzorii. O parte a iluziei rezida în contrastul înșelător dintre teoretic și practic. O altă parte rezida în confuzia parte/întreg în care plauzibilitatea judecății era măsurată cu ajutorul plauzibilității regulilor individuale utilizate pentru a face acea judecată. Și o a treia componentă era o încredere exagerată în analogia dintre om și animal. Odată corectate aceste greșeli, ne vom aștepta ca nivelul general de performanță să fie destul de ridicat, chiar dacă va fi situat sub nivelul optim. În plus, omul va avea zone slabe în care ceva mai mult decât ghinionul sau „zgomotul“ este responsabil pentru erorile sale. Eșecurile cognitive vor fi sistematice, „erori de proiectare“ cum spun inginerii.

Din fericire, există un pendant pozitiv al acestei calificări negative. Omul ar avea zone de forță în care se ridică mai presus de condiția sa modestă. Existența acestor puncte de strălucire poate fi stabilită cu aceleași tipuri de mecanisme care generează slăbiciunile omului.

1. *Caracteristici independente:* Pleiotropia poate fi furnizorul regulilor supernormale de deducție la fel de bine ca și al erorilor. Aici, o regulă de deducție plauzibilă dar inutilă este menținută datorită legăturilor sale cu o caracteristică utilă.

2. *Capacități secundare:* În cazul pleiotropiei, cele două caracteristici sunt legate printr-o cauză comună. Este de asemenea posibil ca o caracteristică inutilă să fie derivată dintr-o caracteristică utilă de mai mare adâncime. Astfel, nu este nevoie

ca toate intuițiile noastre modale să fie folosite. Credințe modale inutile dar plauzibile pot să apară în calitate de capacități secundare care merită să fie păstrate. Informația că este posibil ca un cub să fie transformat într-un cub mai mic, poate fi un produs nepractic al unor capacități practice. Aceste însușiri ale judecății sunt pandanți ai „înzeestrărilor naturale“ (fenomene interesante la care suntem, din fericire, expuși).

Unele produse secundare sunt întregi inutile, alcătuiți din părți utile. Selecția naturală poate favoriza, de asemenea, caracteristici inutile individual, dar utile în mod colectiv. George C. Williams sugera că „însușirile mentale superioare pot fi un efect accidental al selecției capacităților de înțelegere și memorizare a instrucțiunilor verbale din copilărie“<sup>23</sup>. Copii care urmăresc instrucțiunile sunt mai puțin expuși morții prin accidente. Oricum, capacitățile componente necesare acestei caracteristici pot fi redirecționate spre rezolvarea unor sarcini intelectuale mai dificile.

3. *Capacități construite* : Ceea ce se petrece întâmplător poate fi făcut să se petreacă intenționat. Astfel de „efecte de poziție“ pot fi cultivate deliberat. Să considerăm modul în care oamenii inteligenți fac ca aparatele de măsură să răspundă la întrebări pentru care ele nu au fost proiectate să răspundă. De exemplu, excursioniștii ingenioși își folosesc ceasurile drept busole. Țin ceasul pe un plan orizontal, cu indicatorul orar îndreptat spre soare. Sudul se află spre indicatorul orar și ora 12. Deci, dacă este ora 4, Sudul se află în direcția orei 2. Oamenii utilizează metode de raționare în acest mod indirect, pentru a prevedea comportamentul altor oameni. Informația că celălalt acționează pe baza unei metode de raționare îți permite să-i prezici comportamentul, reproducând modul său de raționare. Reproducerea are loc prin substituirea dorințelor și convingerilor tale cu ale sale. Tehnica poate fi extinsă la persoane mai curând neasemănătoare ție (copiii și străinii) și chiar la anumite animale. Într-adevăr, această tehnică poate fi împinsă dincolo de domeniul literaturii. Procesele naturale sunt guvernate de principii asemănătoare cu principiile de luare a deciziilor: apa alege calea minimei rezistențe, plantele caută lumina Soarelui, sexele opuse se atrag. Viziunea teleologică asupra naturii ne permite să ne testăm puterile.

Aceste exemple de triangulație mentală sprijină o distincție între experimente mentale directe și indirecte. Încercările și erorile de substituție sunt experimente de încercare și eroare pe cât de directe, pe atât de simple. Ele se bazează pe un test direct al erorii. Într-un experiment indirect, se utilizează un test indirect. Prin urmare, în cazul experimentului mental, se pot obține progrese chiar dacă lipsesc intuițiile direct relevante în problemă. De exemplu, noi adesea abordăm o întrebare punând o altă întrebare, care exagerează semnificația problemei puse. Să considerăm o ghicitoare, formulată cu intenția de a înlesni înțelegerea legii lui Arhimede despre corpurile care plutesc. Principiul spune că un obiect care plutește dislocă un volum de apă egal cu masa obiectului respectiv. Deci, masa unei monede poate fi determinată punând-o pe o plută, într-un recipient și măsurând masa apei dislocuite. Ce se întâmplă cu nivelul apei dacă moneda este aruncată în recipient? Aici ne este de ajutor să ne imaginăm că moneda cântărește o tonă. Atunci ea dislocă multă apă plutind pe suportul său. Dar odată ce este scufundată în recipient, moneda dislocă doar atâta apă cât să compenseze *volumul* său. Deci, răspunsul este că nivelul apei va scădea. Strategia opusă este de a transforma o cantitate foarte mare într-o cantitate obișnuită. De exemplu, *Dr. Tomkins în Țara minunilor*, al lui George Gamow se află într-o lume în care viteza luminii este de trei mile pe oră, pentru a se evidenția efectele relativiste.

4. *Reguli cu cost sporit* : O regulă nouă care înlocuiește o regulă veche, nu trebuie să fie întotdeauna mai puțin costisitoare, inferioară regulii vechi. Uneori, regula veche permite aceeași performanță sau chiar o performanță superioară pentru același cost sau chiar pentru un cost mai redus. O precizie surprinzătoare poate fi obținută prin măsurarea unor proprietăți aparent irelevante. Unii fluturi măsoară umiditatea după indicațiile *vizuale* ale ciclului zi-noapte al luminii. Plantele se pregătesc pentru frigul iernii măsurând diminuarea intensității luminii diurne – iar nu media variației temperaturii.

5. *Absența înlocuitorilor necostisitori* : Uneori trebuie să mergi cu clasa I, deoarece clasa a II-a nu este disponibilă. Atunci când se întâmplă așa, subiectul cunoscător dispune de un surplus stabil de plauzibilitate.

Distribuția nesistematică a câștigurilor la punctele 1 – 5 va introduce diferențe de la cer la pământ în performanțele noastre. Noam Chomsky specula faptul că aceste neregularități sunt răspunzătoare pentru unele progrese neuniforme în știință<sup>24</sup>. Probabil că unele domenii prosperă, în timp ce altele stagnează, din cauză că oamenii au înzestrări biologice diferite.

Lecția pentru experimentul mental este că pot exista mari variații de plauzibilitate. Experimentul mental poate să aibă succes într-un domeniu și să dea greș într-un alt domeniu învecinat. Insuccesele pot fi tot atât de dramatice ca și cele resimțite de un afazic care poate citi, dar nu poate scrie. Descoperirea că această variabilitate este corelată cu domeniile ar putea produce temeiuri pentru a limita experimentul mental la unele domenii și a-i extinde rolul în raport cu alte domenii. Acest fapt ar respinge neutralitatea metodei în raport cu obiectul. Experimentul obișnuit a putut scăpa acestor restricții întrucât astronomia și paleontologia sunt, ambele, domenii non-experimentale.

## V. FUNDAMENTUL BIOLOGIC

Biologia din spatele intuițiilor noastre modale stabilește un fundament instructiv al performanțelor așteptate de la experimentul mental. Când studiem intuițiile modale ar trebui să ținem cont de faptul că ele sunt orientate către luarea de decizii concrete. Întrucât experimentele mentale și-au câștigat statutul sugerând precauții și ocazii favorabile vânătorilor-culegătorilor, intuițiile modale ar funcționa excelent în raport cu problemele pe care aceste ființe umane le aveau de rezolvat.

Astfel, intuițiile ameliorează buna cunoaștere, profitul, concretețea, manevrabilitatea, simplitatea și vizualizarea. Intuițiile s-ar clătina atunci când împrejurările ar fi fără precedent, monotone sau complexe.

Un experiment mental reușit trebuie să ne permită să ne testăm puterile. Domeniile deliberative (etică, dreptul, științele

economice) sunt avantajate, deoarece ipotezele lor au deja un aspect decizional. Caracterul deliberativ nu este defel necesar, dar este un pas în direcția justă. Exercițiile din manualele de fizică înfățișează unele aprecieri ale problemelor în discuție. Ele dau adesea o nuanță practică ipoteticului, punând o problemă de pierdere, optimizare sau strategie. În orice caz, aproape orice experiment mental este ameliorat prin cantonarea sa în domeniul intuițiilor mai plauzibile.

Din nefericire, adesea ne lipsește orice intuiție directă în materie, în special în problemele înalt teoretice. Experimentatorul obișnuit se confruntă cu același tip de dificultate atunci când nici o percepție directă nu poate duce la soluția problemei. În ambele cazuri, singura soluție este abordarea indirectă. Construcția inteligentă a unui scenariu poate pune în lumină o altă intuiție care la început a putut să pară irelevantă.

Fundamentul biologic ne conduce la erori la fel ca și la adevăruri. Cel puțin la prima vedere, tabloul evoluționist sugerează că noi am avea o capacitate excelentă de raționare. Căci ilogicitatea produce ineficiență, iar ineficiența este înlăturată de selecția naturală. Dar psihologia experimentală a pus în evidență existența unui mare quantum de incompetență<sup>25</sup>. Oamenii obișnuiți sunt surprinzător de slab înzestrați în privința deducției și a raționamentelor statistice. Evident, Mama Natură a preferat ca noi să abordăm sarcinile de raționare formală bazându-se în principal pe memorie și analogie. Chiar și statisticienii profesioniști tind să alunece în acest stil primitiv de gândire atunci când expertizele lor sunt inoperante.

Fundamentului biologic i se subestimează performanțele. Inaccesibilitatea la informații directe, privind organele noastre interne sau stelele îndepărtate, nu au împiedicat progresul în fiziologie și astronomie. Experimentele mentale ale lui Einstein, elaborate în sprijinul teoriei relativității, au în vedere viteze foarte mari. Ipoteza superconductivității este legată de temperaturi foarte joase și de fenomenul ciudat al electricității. Într-adevăr, fundamentul biologic pare a fi condamnat la o subestimare sistematică, deoarece el omite aspectul social al științei. Multe instituții stimulează difuzarea și creșterea cunoașterii. Infrastructura din spatele științei ne permite să depășim multe obsta-

cole care par insurmontabile dintr-un punct de vedere biologic îngust. De exemplu, calculatoarele și diviziunea intelectuală a muncii au diminuat caracterul problematic anterior al complexității.

Subestimările sunt reale. Dar să notăm că greșelile sunt cele pe care avem a le face. În loc să fie impedimente în cunoaștere, ele sugerează distincții interesante. De exemplu, succesul surprinzător al experimentelor mentale în legătură cu fenomenele neobișnuite indică până unde putem fi elitiști. Egalitarismul este în regulă atunci când controversele privesc fenomene obișnuite, dar numai intuițiile experților au șanse în condiții neobișnuite. Căci numai experții și-au rafinat aceste intuiții în lumina teoriilor și experimentelor care pot avea legătură cu un anumit fenomen.

Predicțiile greșite ale fundamentului biologic sugerează, de asemenea, procese interesante. Să considerăm din nou previziunile sumbre despre implauzibilitatea universală a experimentelor mentale care vizează fenomene neobișnuite. Aceste previziuni sunt respinse de existența experimentelor mentale reușite în teoria relativității, mecanica cuantică și cercetările privind superconductivitatea. Totuși noi am continua să credem că rata succesului este mult mai scăzută decât pare și astfel să bănuim existența unui corp remarcabil de mare de experimente mentale eșuate în aceste domenii. Această bănuială este înlăturată de istoria teoriei relativității. Bunul simț nu este opac în legătură cu ceea ce se întâmplă la viteze foarte mari. El extrapolează cu ușurință mărimile medii. Astfel, oponentii relativității nu întâmpină nici o dificultate în a colecționa experimente mentale respinse. Dar ei greșesc cu atât mai mult cu cât sunt mai numeroși. Experimentele mentale eșuate sunt mai puțin importante, deoarece ele, odată expuse, încetează a mai fi menționate. Astfel, o parte a succesului nostru aparent obținut cu ajutorul experimentelor mentale este o mostră de iluzie. Acest reziduu al barierei biologice se extinde și la experimentele obișnuite. Studenții care învățau experimente din manuale, supraestimau puterea de decizie, claritatea și plauzibilitatea metodei experimentale.

Ar trebui, de asemenea, să fim pregătiți să pătrundem în stocul nostru de explicații deja acceptate. Probabil că unele experimente mentale doar par să fie reușite. Puterea lor de convingere

poate fi explicată ca fiind mai curând un efect retoric sau o confuzie între contextul descoperirii și contextul justificării. Oricum, scoaterea din cauză a multor cazuri aparent clare este o tactică greșită. Paradigmele noastre inițiale ar trebui să primească, pe ici, pe colo, un tratament mai aspru.

Deci, teoria evoluționistă a experimentului mental nu este în nici un caz o teorie completă. Ea necesită amendamente la nivelul științelor sociale pentru a corecta pesimismul său epistemologic rezidual. Din fericire, această tendință, precum și incompletitudinea abordării evoluționiste nu o împiedică pe aceasta să sugereze întrebări fecunde și răspunsuri corecte.

## NOTE

<sup>1</sup> „The Nature of Laws“ în *Canadian Journal of Philosophy*; vol. 7, 1977, 76.

<sup>2</sup> Contrastul este expus în corespondența lui Mersenne, din 1931 – 1932, cu Jean Rey. Cuprinsul acestui schimb de scrisori este relatat în lucrarea lui Pierre Duhem, *The Aim and Structure of Physical Theory*, trad. de Philip P. Weiner (New York: Atheneum, 1974).

<sup>3</sup> *The Nature of the Physical World* (1928) p. 267.

<sup>4</sup> *The Genetical Theory of Natural Selection* (Oxford: Oxford University Press, 1930) pp. VII – IX.

<sup>5</sup> J. Loschmidt, „Ueber den Zustand und des Warmegleichgewichtes eines System von Korpern“, *Akademie der Wissenschaften, Wien. Mathematisch-Naturwissenschaftliche Klasse. Sitzungsberichte*, 59, Abth. 2 (1876) pp. 137 – 9.

<sup>6</sup> Ludwig Boltzmann, „Bemerkungen über einige Probleme der mechanischen Wärmetheorie“ (1877), *Wissenschaftliche Abhandlungen von Ludwig Boltzmann*, ed. F. Hoesenohrl (Leipzig, 1909) vol. II, 120.

<sup>7</sup> James Robert Brown apără o poziție godeliană în „Thought Experiments since the Scientific Revolution“, *International Studies in the Philosophy of Science*, vol. 1, no. 1 (September 1986) pp. 1 – 15; p. 13.

<sup>8</sup> William Whewell, fragmente nepublicate din *Of the Plurality of Worlds* (Londra, 1853) citate în R. E. Butts, „Necessary Truth in Whewell's Theory of Science“, *American Philosophical Quarterly* vol. 2 (1965) pp. 161 – 81; p. 179 fn. 41.

<sup>9</sup> Ronald Curtis analizează epistemologia lui Darwin în „Darwin as an Epistemologist“, *Annals of Science*, vol. 44 (1987) pp. 379 – 408.

<sup>10</sup> Charles Darwin, „My Notebook“, în *Metaphysics, Materialism, and the Evolution of Mind*, ed. Paul H. Barnett (University of Chicago, 1974) p. 30.



<sup>11</sup> Ch. Darwin, „Essays on Theology and Natural Selection“, în *Metaphysics, Materialism and the Evolution of Mind*, p. 160.

<sup>12</sup> W.K. Clifford, „Philosophy of the Pure Sciences“, în *Essays and Lectures*, L. Stephen, Fr. Pollock, Londra, Macmillan, 1901, pp. 335–336.

<sup>13</sup> Darwin către Wallace, *More Letters of Charles Darwin*, ed. F. Darwin și A. Seward (New York, 1903) vol. 1, p. 395.

<sup>14</sup> Cea mai completă examinare a problemei o întâlnim în lucrarea lui Mach: „On Thought Experiments“, în *Knowledge and Error* (Dordrecht: D. Reidel, 1976), apărută în 1905. Dar o serie de comentarii ale sale apar în lucrări anterioare.

<sup>15</sup> „Geological Climates and the Origin of Species“, în *Quarterly Review*, (1869).

<sup>16</sup> „My Notebook“, în *Metaphysics, Materialism and the Evolution of Mind*, ed. Paul H. Barnett (University of Chicago Press), 1980, p. 78. De asemenea, *The descent of man* (New York: Hurst, 1871), p. 49.

<sup>17</sup> „Kant's Doctrine of the *A Priori* in the Light of Contemporary Biology“, *General Systems*, vol. VII, ed. L. von Bertalanffy and A. Rappoport (Ann Arbor: Society for General Systems Research, 1962), pp. 23 – 35; p. 25. Aceasta este o traducere dintr-un articol apărut în *Blatte fur Deutsche Philosophie*, vol. 15 (1941) pp. 94 – 125.

<sup>18</sup> „Evolutionary Epistemology“, în *The Philosophy of Karl Popper*, ed. Paul Arthur Schilpp (La Salle, Illinois: Open Court, 1974), pp. 413 – 463, p. 447.

<sup>19</sup> *Behind the Mirror*, trad. R. Taylor (Londra, Methuen, 1977, p. 147).

<sup>20</sup> Ch. Darwin, *The Descent of Man*, (New York, Hurst, 1871 p. 99).

<sup>21</sup> Aceasta este o versiune ușor adaptată a argumentului pe care Stephen Stich l-a utilizat pentru reconstrucția rațională a argumentării lui Daniel Dennett. Vezi p. 123 din lucrarea lui Stich: „Could Man be an Irrational Animal“, *Synthese*, vol. 64 (1985), pp. 115 – 135.

<sup>22</sup> J. Garcia, B. K. McGowan, și K. F. Green, „Biological Constraints on Conditioning“, în *Classical Conditioning II*, ed. A. H. Black și W. F. Prokasy (New York: Appleton-Century-Crofts, 1972).

<sup>23</sup> *Adaptation and Natural Selection* (Princeton University Press, 1966, p. 15).

<sup>24</sup> Vezi capitolul al patrulea al lucrării sale *Reflections on Language* (New York: Pantheon, 1975).

<sup>25</sup> Un bun exemplu de asemenea cercetare este conținut în lucrarea *Judgement under Uncertainty*, ed. D. Kahneman, P. Slovic și A. Tversky (Cambridge University Press, 1982).

Traducere Dumitru GHEORGHIU

# EXPERIMENTUL MENTAL DIN PERSPECTIVA PSIHOLOGICĂ

Ion MÂNZAT

(Universitatea din București)

## 1. EXPERIMENTUL POZITIV ÎN PSIHOLOGIE

Claude Bernard (1958) a încercat o definiție și o descriere a conceptului de *experiment științific* insistând asupra operației centrale de disociere a factorilor și asupra discriminării între factori reali și aparenti. Experimentul științific presupune elaborarea unei strategii de explorare și efectuarea unor măsurători. Metoda experimentală a lui Cl. Bernard procedează prin variațiuni concomitente. Se reproduce de sute sau de mii de ori același experiment, dar variind de fiecare dată unul din factori. Se notează rezultatele obținute și se desprind regularitățile. Este o metodă care și-a dovedit eficacitatea, dar nu este infailibilă și nici aplicabilă în toate domeniile sau problemele.

*Experimentul psihologic* (clinic, natural, de laborator etc.) se distinge prin preocuparea de a discrimina, pe căi indirecte, subiectivul de obiectiv nu numai sub aspect cantitativ ci și calitativ. Experimentul pozitiv (mai ales cel de laborator, cel mai frecvent folosit) verifică ipoteze, măsoară rezultate, vizualizează prin grafice, caută să convingă prin formule statistice. De când s-a desprins de filosofie (perioada 1860–1880) psihologia și-a câștigat prestigiul de știință experimentală și a dobândit o mai mare credibilitate și prin utilizarea sistematică a experimentului de laborator. Dar nu toate procesele și fenomenele psihice pot fi studiate numai pe această cale. Avem în vedere complexe combinatorici imaginative, intuițiile, motivația, voința, atitudinea, creația, speranța, credința, transcenderea etc.

Claude Bernard declara: „Un om care descoperă faptul cel mai simplu aduce mai multe servicii omenirii decât cel mai mare filosof din lume“. Această declarație ne apare ca o culme a poziti-

vismului extremist, unilateral și absolutizant. Știința n-ar putea fi decât experimentală (în sens pozitiv). În afara ei nu e posibilă nici o altă cale. Să închidem ușile. Nimeni nu-i va egala vreodată pe giganții care au inventat mașina cu aburi. În acest univers organizat, comprehensibil și condamnat omul trebuie să stea în locul cuvenit de epifenomen. Fără speranță și fără utopie. Karl Jaspers își exprimă dezamăgirea și mirarea: „Este uimitoare pasiunea cu care se cufundă europeanul modern în toate lucrările, chiar în cele altădată disprețuite, numai pentru că sunt empiric reale.

Creatorul din știință și tehnică, artă și economie – ca și magicianul, sub ochii logicianului cartezian – trece prin *bariere* datorită unui fenomen analog celui numit de fizicieni „efectul de tunel“. O aspirație magică îl atrage și îl domină. Vrea să vadă prin ziduri, să ajungă pe Marte, să capteze fulgerul, să studieze străfundurile oceanului planetar, să citească gândurile celorlalți etc. Caută să prindă universul în flagrant delict de umblat cu ascunzișuri. Vrea să sfarme „frânele transcendente“ din calea misterelor (L. Blaga, 1993): Vrea să decodifice simbolurile arhetipale și să prevadă viitorul.

## **2. EXPERIMENTUL MENTAL – PREMISĂ ȘI REZULTANTĂ A EXPERIMENTULUI POZITIV**

Cercetătorii cu vocație, înainte de a-și realiza programul sau modelul experimental își construiesc *proiectul mintal* al întregului demers pe un fond de tensiune intelectuală creată prin confruntarea logică a ipotezelor și contraipotezelor, a strategiilor euristice și algoritmice. Întregul experiment se derulează întâi în mintea cercetătorului. După încheierea experimentului extern cercetătorul trebuie să conceapă mintal modul cel mai adecvat de interpretare și prezentare a rezultatelor și să se concentreze asupra esențelor și regularităților din care poate desprinde un nou concept sau o nouă teorie.

Experimentul psihologic este un astfel de *proces* în care sunt implicate deopotrivă operații obiectuale, operații intelectuale și operații de gradul doi („operații asupra operațiilor” – J. Piaget, 1965).

Psihismul uman nu este doar o *reflectare* subiectivă a lumii obiective, ci este și *construcție-creație*, o „creștere din interior” (John Dewey). Multe idei, trăiri izvorăsc în mod misterios din psihism. J. Piaget a arătat primul că pe lângă operații ale inteligenței care sunt o interiorizare a acțiunilor obiectuale (operațiile concrete) există și operații formale sau „propoziționale” sau operații de gradul doi și chiar de gradul trei fără vreo sursă evidentă în realul empiric. Așa, de exemplu, se poate dezvolta „abstractizarea reflectantă” (abstractizarea abstractizărilor), generalizarea generalizărilor, sinteza sintezelor etc.

Iată deci că inteligența umană este înzestrată cu capacitățile necesare realizării și trăirii experimentului mental care se configurează ca un *complex operațional* care se derulează și se dezvoltă la nivelul mentalului sub forma confruntării și conjeturării ipotezelor, teoriilor, conceptelor prin transferuri și contra-transferuri ca „extinderi ale câmpurilor operatorii”. Am reușit să surprindem, în cadrul gândirii științifice, transferuri directe (extrapolări), transferuri inverse (recurențe), transferuri complementare și structurale, transferuri de gradul doi („transferuri ale transferurilor” – I. Mânzat, 1988).

Experimentul mental presupune din partea cercetătorului un efort de *introspecție* care trebuie să-l ajute să se dedubleze: o dată este inițiatorul și realizatorul unui experiment și totodată este verificadorul care controlează, contrazice și rectifică. Experimentatorul este succesiv sau simultan proiectant și executant, actor și spectator, judecat și judecător.

Experimentul mental este un *modus operandi mental* care nu exclude întotdeauna „realul empiric”, ci îl imaginează *esențializat*, adică purificat, curățat de detalii, artefacte și elemente nesemnificative.

În știință găsim nenumărate descoperiri care sunt rodul experimentelor mentale. În fizică – teoria relativității restrânse și generalizate (A. Einstein), principiul complementarității (Niels Bohr), principiul nonseparabilității particulelor (Bernard

D'Espagnat) ș.a. În psihologie – conștiința ca sinteză creatoare (Wilhelm Wundt), inconștientul transpersonal care conține arhetipurile (Carl Gustav Jung), principiul sincronicității sau „al relațiilor acauzale prin coincidență semnificantă” (C.G. Jung și Wolfgang Pauli) etc. În aceste cazuri experimentul pozitiv este neputincios și perturbator.

Gândirea matematică începând cu notele din scrisoarea tânărului Evariste Galois (1811–1832) a descoperit o lume străină omului, care nu corespunde experienței umane obișnuite, universului așa cum îl cunoaște conștiința omului mediu. Logica procedând prin *da* sau *nu* este înlocuită cu o superlogică funcționând prin *da și nu*. Această superlogică nu este de domeniul raționamentului, ci al *intuiției*. Lucrările lui Galois constituie punctul de plecare în studiul algebrei moderne și în teoria actuală a funcțiilor algebrice. Uneori gândirea științifică a unei epoci nu este pregătită pentru a recepționa și înțelege o invenție sau o descoperire. Astfel, ideile revoluționare ale lui Galois au fost respinse de Academia Franceză de Științe ca „neinteligibile” și numai 15–20 de ani mai târziu au început să fie acceptate și înțelese. Galois și-a elaborat teoria la vârsta de 20 de ani, aceasta fiind consemnată într-o scrisoare adresată unui prieten, în noaptea premergătoare duelului în care a fost ucis. Se pare că el este cel mai tânăr savant din istorie. Academicienii francezi din 1831 au încercat să „descifreze” notițele lui Galois apelând la raționamentul matematic, riguros și sistematic dar neputincios în acest caz. Intuiția genială aștepta să comunice tot cu intuiția.

Se vorbește astăzi de existența unui *hiperspațiu* (spațiu transfizic) răspunzător de producerea unor evenimente și fenomene exotice, a unui spațiu cu 4, 5, 6, 11 sau  $n$  dimensiuni în care pulsează energii transfizice, manifestate sau nu în sistemele vii, cu proprietăți distincte de cele ale energiilor fizice, de o rară subtilitate a materiei, cu o *sincronicitate acauzală*, dezbrăcată de suportul temporal, pusă sub semnul asociativității între universul supra-senzorial și cel fizic, un fenomen de impregnare sau aderență între forțele cosmice ori telurice și anumite persoane.

Dar tentația de a rămâne la „continuumul material”, în planul exclusiv al experimentului fizic (eliminând componenta spirituală a universului ca neștiințifică), își refuză deocamdată orice certitudine

în planul cunoașterii ezotericului. Psihologia paranormalului rămâne încă în opoziție conflictuală față de cercetarea bazată pe experimentul pozitiv. Prea marea încredere pe care o au oamenii în știința experimentală și în cunoașterea pozitivistă poate bloca însăși conștiința integrării omului în cosmos, poate conduce la o ecranare a intuiției, la tocirea simțului universalității, la limitarea libertății spiritului. Multe din marile cuceriri ale gândirii științifice au fost posibile tocmai grație unei suspendări a raționamentului riguros al experimentului obiectual și prin încercarea de a asculta vocea intuiției creatoare și a paradoxului. Astfel, numai prin experiment mental și intuiție cognitivă au putut ajunge David Bohm și Jean Pierre Vigier (în 1954) la ipoteza existenței unui „mediu subcuantic“, ascuns și subiacent: o particulă este mereu în contact cu un mediu ascuns care corespunde unui nivel mai profund al realității fizice decât nivelul cuantic, mediu cu care toate particulele de la nivel cuantic s-ar afla permanent în interacțiune.

Deseobit, experimentul mental este atât premisa cât și rezultanta experimentului obiectual.

### **3. ANALIZA COMPARATIVĂ A CELOR DOUĂ TIPURI DE EXPERIMENT**

Experimentul pozitiv este programatic organizat, riguros controlabil și verificabil statistic. El poate fi repetat oricât de des. Rezultatele sale sunt convingătoare de la început. Experimentul mental este impregnat cu ipoteze și imagini, cu scheme abstracte, concepte figurale și credințe ale cercetătorului. Este mai liber și mai flexibil, dar poate ușor să decadă în ficțiune și fantasmatic. Experimentul mental este liber de pericolul erorilor de observație și al erorilor de instrument, dar nu este eliberat de erorile și aproximările imaginației creatoare.

Sigmund Freud (1893) observa: „Când viața ne pliază sub severa sa disciplină simțim în noi o rezistență contra inexorabilului,

monotoniei gândirii, exigențelor probelor realității. Pentru că ea ne privează de multe prilejuri de plăcere, rațiunea devine un inamic de sub jugul căruia ne smulgem cu bucurie, cel puțin temporar, abandonându-ne seducțiilor neraționalului“.

Experimentul pozitiv utilizează frecvent analiza logică și controlul *rațional*. Experimentul mental este condus de către imaginație care dacă nu cooperează cu gândirea poate degenera în fantastic și neplauzibil. Experimentul mental acceptă însă și neraționalul, îmbinând adesea raționalul cu neraționalul, sub-raționalul cu suprraționalul avansând până la un nivel în care ele se combină complementar în procesul complex al intuiției creatoare.

Cele două tipuri de experimente, au fiecare atât avantaje cât și dezavantaje. Avantajele lor însă pot deveni complementare.

Astăzi experimentul științific ia o turnură sinuoasă care avansează de la exotic (adică de la negarea statutului său prezent) prin idee, măsură, inducție, transducție, proiect mental, intuiție, relație de coordonare, complementarizare spre o integrare finală care se aseamănă foarte mult cu relația parte-întreg din gândirea hindusă, cu relația complementară *yin* versus *yang*, din taoismul chinez sau cu relația unu – multiplu din filosofia plotiniană. Fizica cuantică devine o împăcare de contrarii care sunt puse să conlucreze: corpuscul și undă, formă și vibrație, haos și cosmos, manifest și ascuns, nimic și ceva. Imaginea holografică a universului (dar și a creierului!) apare pe urmele monadei leibniziene, dar și ale Aleph-ului cabalistic. Multe teorii și principii moderne sunt actualizări ale unor ontologii arhaice.

Faptul că știința nu poate ieși din criza propriei sale supra-producții decât prin valorizarea unei „relații de incertitudine“ (Werner Heisenberg), a unei logici a contradictoriului (Gödel), a integrării contrariilor într-o „a treia stare“ care să echilibreze balansul potențialitate/actualizare – omogenizare/eterogenizare (Șt. Lupasco) este semnificativ pentru necesitatea limitării căii pozitiviste a cunoașterii. Paradoxul spațiu-timp, ca linie continuă a universului, paradoxul marii reunificări (G.U.T.) a celor patru forțe primordiale ori cel al viziunii holografice a universului (Davis Bohm), sau teoria câmpurilor sau a supercorzilor vibratorii unde haosul capătă, alternativ, aceeași valoare cu cosmosul sunt

încercări de a ieși din perimetrul noofag al exactității tehnico-științifice și de a regăsi sensurile originare ale lumii (Vasile Avram, 1994, p. 267). Numai gândirea metafizică și meapsihică poate accede până la descrierea misterelor și esențelor ascunse. Pentru o astfel de gândire „meta – de gradul doi“ nu curgerea ca de râu caracterizează lumea, ci rotirea, spirala, cercul cu centrul peste tot și circumferința nicăieri (Blaise Pascal). Distanța în spațiu și în timp nu se măsoară cu unități divizibile, ci într-o cuprindere globală, cu toate simțurile reunite într-unul singur, în mod sinergic, care s-ar putea numi *simțul ciclopic*.

Parapsihologia a demonstrat că telepatia nu se bazează pe transmițeri prin câmpuri electromagnetice sau alt tip de câmpuri fizice, ci prin câmpuri extrafizice, de natură spirituală, în care distanțele în timp și spațiu nu mai contează. Aceasta înseamnă că psihismul uman este o energie nefizică, cu posibilități nelimitate de propagare. Psihologii și fizicienii au condus investigația lumii până în pragul pulverizării conceptelor fundamentale de materie, energie, timp, spațiu, determinare, până în punctul în care universul „cu cât devine mai de înțeles, cu atât pare mai fără sens“ (Sorin Comoroșan, 1993). După divizarea progresivă a particulelor elementare până la a nu mai fi detectabile decât prin ecuații matematice, a urmat o valorizare existențială a vidului (*vacuum*-ul cuantic care fluctuează și, prin acumulare de energie din sine însuși, explodează și naște substanța), o încercare de „a muta zidul transcendentului“, a anihila deosebirea dintre haos și cosmos, dintre substanță și idee, de a reduce totul la o *supervibrație* ordonată în „supercorzi“ și vegheată de o „superforță“ mai impersonală decât cel mai abstract zeu.

Gândirea științifică tinde adeseori la o schematizare care poate să ne pună în situația de a *mânui lucrurile, dar nu de a trăi lucrurile*. Dar noi, oamenii, când gândim lucrurile le și trăim (Nae Ionescu, 1993, p. 86). Experimentul mental, mai mult decât cel obiectual, ar fi deci *trăirea directă* a realității. Dar această trăire afectiv-motivațională sau volițională este individuală și inefabilă sau necomunicabilă. Pentru a fi împărtășită (întrucât știința poartă asupra generalului, nu asupra individualului) ea trebuie transfigurată și tradusă prin *logică* și prin *intuiție supralogică*.



#### 4. COMPONENTELE PSIHICE ALE EXPERIMENTULUI MENTAL

Pentru realizarea experimentului mental acționează o pluralitate de procese și funcții psihice. În centrul acestora, ca un nucleu de generativitate și dinamicitate se mișcă *intuiția cognitivă*. Psihologii clasici (gestaliștii germani) considerau că intuiția ignoră logica, operațiile intelectului și rațiunea. Credem că mai curând le include, depășindu-le. Intuiția poate apărea ca o cunoaștere misterioasă de origine lăuntrică sau manifestându-se ca un sentiment legat de o viziune clară, concentrată. Demersurile intuitive au loc dincolo de liniaritatea proceselor intelectului și de aceea par suspecte. Pe deasupra, intuițiile mai sunt mijlocite de tăcuta emisferă dreaptă (R. Sperry, 1982). Simbolurile ei, imaginile și metaforele ei trebuie preluate de emisfera stângă și reformulate înainte ca o informație să fie cuprinsă în întregime. Intuiția nu e separată de intelect, ea include intelectul.

În domeniul intuiției cognitive, care deseori este generatorul și stimulatorul experimentului mental, psihologia clasică a remarcat efectul dar n-a putut elucida *procesul* sau l-a expedit, repede și superficial, în abisul inconștientului. Tot ceea ce se bănuiește a ține de inconștient are un mare avantaj: nu mai trebuie explicat și justificat pentru că inconștientul a rămas încă abscons și abisal. Psihologia definește intuiția ca „descoperire bruscă a unei soluții, fără conștiința modului de descoperire” (Paul Popescu-Neveanu, 1978, p. 380). Suntem de acord cu caracterul de explozie, de erupție, de străfulgerare scurtă și neașteptată. Dar acesta nu este decât punctul *terminus* al unui proces. Mulți psihologi nu sunt de acord că intuiția este proces, pentru că aceasta ar însemna o durată cu desfășurarea succesivă a unor etape. Intuiția, după ipoteza noastră (I. Mânzat, 1995), este un proces psihic complex, care se desfășoară atât de rapid încât succesiunea pare simultaneitate. *Intuiția este psihosinergie*: mai întâi se realizează *cooperări sinergice locale* între procesele cognitive; apoi nucleul de generativitate și expansiune devine „imaginația conceptuală” (B. D’Espagnat) ca o complementaritate a gândirii și imaginației; sinergia din câmpul cogniției se extinde rapid în

celelalte câmpuri extracognitive până când toate aceste cooperări sinergice locale comunică între ele și se combină neașteptat izbucnind ca o „erupție vulcanică“, *cooperarea globală* a tuturor proceselor și funcțiilor psihice, conștiente și inconștiente, raționale și iraționale. Acesta este momentul final, în care cercetătorul *percepe această psihosinergie globală ca pe o intuiție*; el remarcă doar emergența, dar îi scapă procesul (incubația). Studiind intuiția, ca energizor al experimentului mental, nu trebuie să excludem iraționalul din câmpul inteligenței umane. Nu este corect să explicăm intuiția degradând-o și coborând-o în inconștient sau subconștient. Din moment ce intuiția este o funcție care depășește clar potențialul obișnuit al conștiinței, este absurd și nedrept să o plasăm dedesubt, în inconștient. Intuiția creatoare este o funcție ascendentă, care se ridică deasupra conștiinței, până la un nivel pe care noi îl numim *transconștiință* – conștiința de vârf, a patra stare de conștiință sau conștiința cea mai performantă care funcționează rareori și nu la toți oamenii (I. Mânzat, 1955). Intuiția poate fi funcția *princeps* a experimentului mintal ca un demers idealizat, esențializat. Procesându-se ca psihosinergie globală intuiția este o *generatrice* a experimentului mental, adică o matrice generatoare de reorganizări și combinații creatoare, care se află în relații de alianță și nu de adversitate cu inteligența. Numai pe baza acestei alianțe care nu exclude și momente de competiție este posibil sinergismul psihic generalizat (I. Mânzat, 1991).

Procesele psihice care acționează în planul experimentului mental nu sunt izolate, ci grupate în combinații și complexe procesuale. Importantă este „raționalizarea imaginativă“ (A.N. Whitehead), o metodă opusă drumului inductivist, empirist; este, de fapt, „experimentul imaginativ“ al științei și filosofiei supranumit și „descriere metafizică“. Prin această metodă investigatorul discernе, într-un anumit domeniu, ceea ce el apreciază că ar reprezenta trăsăturile generale și esențiale, după care formulează categoriile și simbolurile pentru această investigație și încearcă să descopere dacă ele există și în alte domenii. Apoi întregul sistem se confruntă deductiv cu faptele experienței. C. Noica (1993, p. 23) apreciază cel mai mult „nu inductivul, nici deductivul, ci *pro-ductivul* – a duce înainte ceva“.

Un rol deosebit în dinamica experimentului mental îl are „imaginația conceptuală” (B. D’Espagnat) care presupune complementaritatea gândirii și imaginației până la sinergismul lor. Gândirea izolată este prea cenzurată și prea conformistă. Imaginația, fără cooperarea cu gândirea, este prea libertină și poate depăși cadrele raționalului și ale plauzibilului. Leibniz arăta că fenomenele imaginare sunt o expresie a universului intern și un mijloc de îmbogățire a acestuia. După P. Malrieu (1967) imaginația se distinge prin posibilitatea de a descoperi și transforma lumea prin intermediul descoperirilor și transformărilor care intervin în propria personalitate. Imaginația conceptuală este specifică creației științifice (descoperire) și tehnice (invenție). Corelația și cooperarea gândire-imaginație în creația științifică poate fi susținută prin analiza psihologică a relației dintre imagine și concept. J.P. Sartre (1940) nu vede o deosebire de principiu între cele două entități. El scrie: „Există pentru concept două moduri de apariție: ca pură gândire pe teren reflexiv, iar pe teren nereflexiv, ca imagine”. În structura imaginii pot intra judecăți pe care Sartre le numește „aserțiuni imaginative”. Paul Popescu-Neveanu (1977, p. 375) arată că în timp ce conceptul e structura tare a intelectului, imaginea este o structură slabă, dar mai bogată. În cadrul creației științifico-tehnice iluminarea se reduce la un singur moment; mai multe momente de inspirație sunt presărate de-a lungul întregului proces creator, în fiecare etapă sau secvență a sa (Sidney Parnes, 1967). „Combinatorica imagistică are un câmp mai larg decât combinatoricile reglementate logic” (Paul Popescu-Neveanu, 1977, p. 375). A. Einstein și N. Bohr în fizică, S. Freud și C.G. Jung în psihologie au săvârșit un efort de „imaginație conceptuală”. Ei au imaginat un real cu totul diferit de datele rațional admise. Pornind de la această proiecție imaginativă au stabilit ansambluri de fapte pe care experiența empirică le-a verificat. H. Bergson (1994, p. 146) arăta că adevărata înțelegere consistă într-o mișcare a spiritului care oscilează între percepții sau imagini, pe de o parte și *semnificația* lor, pe de alta. A experimenta intelectual înseamnă a purta o aceeași reprezentare sau complex imagistic prin diferitele planuri ale conștiinței, într-o direcție care merge de la abstract la concret, de la schemă la imagine.

Fritjof Capra (1983) militează pentru integrarea conceptului de sănătate spirituală, pe de o parte, cu achizițiile științei de până acum și pe de altă parte, cu formele transpersonale ale percepției și experimentului. I se pare deplasată rezumarea la „forma carteziană” a experimentului. Omul de tip cartezian „stă în afară” și premisa sa, precum și rațiunea sa de a fi e „starea în afară”, *ek-sistența* alienată. Tipul acesta de om vedește apoi infirmitatea de a nu înțelege decât ceea ce stă în legătură cu percepția grosieră și cu reprezentarea spațio-temporală. Pentru „realitatea după realitate”, pentru experiența transpersonală, pentru percepția extrasenzorială el nu dispune de valențe comprehensive, nici măcar într-o formă rudimentară. Fr. Capra crede că adevărata sănătate psihică cere o echilibrare a celor două forme de cunoaștere, adoptarea unui nou „stil cognitiv”, un „new mind-scape” (M. Maruyama) caracterizat prin dăruire față de clipă și profundă sesizare a dimensiunii spirituale a existenței.

Experimentul mental nu este doar o modalitate complexă de cunoaștere, întrucât dramatismul proceselor și devenirilor sale relevă și participarea unor *procese extracognitive*: motive intrinseci (curiozitate epistemică, interes cognitiv, atracție spre mister) ca adevărate „cauze interne”, emoții și sentimente intelectuale, capacități de concentrare ale atenției și voinței. Experimentul mental nu este o formă de cunoaștere, ci o *aptitudine* care se poate exprima ca o *capacitate* a subiectului.

## 5. IDEALIZAREA – UN COMPLEX OPERAȚIONAL SINERGIC

*Constructul* se referă la conceptele științifice care se realizează prin idealizare și care se bazează pe abstractizări și generalizări de gradul doi (abstractizări ale abstractizărilor, generalizări ale generalizărilor). După Mario Bunge, constructele sunt *obiecte ideale* construite și experimentate mental. După cum demonstrează G. Uscătescu (1987), conceptele științei sunt *modele mentale*. Această construcție mentală a conceptului se realizează

prin activitatea de idealizare, un complex operațional în care operațiile mentale izvorăsc tot din operații. Gh. Enescu (1973) afirmă că știința nu se poate lipsi de „premise metafizice”: fiind limita ideală a unei tendințe reale, limita la care poate ajunge abstractizarea *obiectului ideal* nu mai este o simplă abstracție... Dacă în abstracțiile simple accentul este pus pe „reținerea unor însușiri”, în cazul idealizării accentul este pus pe „neglijarea unor însușiri”. Conceptele teoretice cu un înalt grad de idealizare sunt numite și *concepte ideale* și pot avea trei semnificații diferite: (a) o semnificație *rațională* (ca „limită ideală”); (b) o semnificație *senzorială* (ca „limită de percepție”, de fapt limitele posibilității noastre de percepere); (c) o semnificație *practică* (limita permisă a rațiunii practice).

Fizica modernă este prima știință care ne-a obligat să luăm în serios problema idealizării. Aceasta se explică prin aceea că fizica, utilizând un aparat matematic foarte sofisticat are de-a face cu „abstracții” de cel mai înalt ordin, iar, pe de altă parte, este mereu obligată să confrunte abstracțiile ei cu realitatea fizică, experimentală. Într-o situație similară se află și psihologia care, deși operează cu abstracții extreme, fenomenele psihice fiind de natură ideală, încearcă să convingă prin cercetări experimentale, de laborator.

Idealizarea este „capacitatea gândirii noastre de a construi așa-zise „obiecte ideale”, adică astfel de concepte în care anumite procese și tendințe reale sunt duse în planul gândirii la limită” (Gh. Enescu, 1973). Astfel, pornind de la faptul real că există corpuri de dimensiuni „din ce în ce mai mici”, gândirea științifică ajunge la ideea „corpului fără dimensiuni”, confirmată de existența unor „imperceptibile” (de dispariția dimensiunilor din câmpul nostru senzorial). Pe de altă parte, procesul gândirii abstracte ne duce la ideea „inutilității”, în anumite operații intelectuale, a însușirilor spațiale. Gândirea științifică ne sugerează posibilitatea de a neglija dimensiunile și timpul astronomic în anumite operații experimentale, în sensul că cercetătorul nu este nevoit să le ia în considerație când cercetează. Toate aceste caracteristici aparțin experimentului mental prin care cercetătorul corelează diversele concepte științifice și le organizează în sisteme conceptuale sau teorii științifice. Mario Bunge (1984) afirmă: „Cu cât este mai

apropiat de experiență, cu atât mai puțin apt este un concept de a intra în axiomele unei teorii... dacă dorim să explicăm experiența trebuie să *ne ridicăm deasupra ei*, prin analizarea în termeni nonexperiențiali“. De aceea, gândirea științifică nu este obligată să rămână permanent în legătură cu realitatea fizică, să asigure valoare referențială tuturor termenilor ei. Aceasta ar frâna dezvoltarea științei. J.P. Vigiér are curajul să declare: „avem întrutotul dreptul de a introduce în fizică lucruri care nu se observă“. Iar W. Heisenberg observa că, în cazul unor concepte ale fizicii teoretice, „relația acestora cu experiența sensibilă se recunoaște foarte greu“.

Conceptul științific este o construcție *ideală* nu numai pentru că aparține sferei ideilor ci și pentru că presupune o anumită *perfectiune*; de caracterul idealizat al construcției conceptuale este legată și ideea de *puritate*, de omogenitate internă. Și întrucât conceptele științifice moderne nu sunt izolate și nici precis conturate, fiind doar „niște penumbre în mișcare“ (Șt. Lupasco) prin experimente mentale se pot întrevădea întrepătrunderile și fuziunile conceptuale.

*Experimentele ideale* care neglijează detaliile și artefactele, imperfecțiunile și „impuritățile“ au avut un mare rol în istoria gândirii științifice dacă admitem că există o parte foarte abstractă numită *formală* a gândirii care modelează nu structuri reale, ci structuri de gândire, fiind compusă din *modele de modele*, din *operații asupra operațiilor* care se referă nu la realul empiric ci la *posibil*. Construcțiile mentale sunt adesea mai reale decât acțiunile obiectuale. Cunoașterea formală se supune altor criterii de evaluare și altor condiții de constructivitate și funcționalitate decât cunoașterea factuală.

Lucian Blaga (L. Sofonea și N. Ionescu-Pallas, 1987, p. 275) observa că știința lucrează și cu *inobservabile* iar ceea ce este observabil rămâne neinteligibil dacă nu este gândit prin modele turnate în modele inteligibile. Blaga dă unele exemple de „obiecte ideale“ obținute printr-un proces de maximă idealizare: conceptele de „punct“ sau „plan“ din geometrie. Punctul este definit ca ceva „spațial“, „fără extensiune“, conceptul fiind „cu un capăt în concret și cu celălalt în abstract“.

Gândirea științifică are o dublă funcționalitate: a) descoperă *relieful faptic* (obiectual, exterior); b) inventează independent, prin idealizare, *raționalul nefaptic* (ne-obiectual, mental). Uneori gândirea *concepe* fără să fi perceput anterior, este „o creștere din interior“ (J. Dewey). De aceea, este posibil ca uneori obiectele raționale, create mental, să *nu* aibă corespondențe factuale. De exemplu, plăsmuiri imaginative ca eterul, curbe-suprafețe imaginate, funcții și functori, simboluri arhetipale, operații atemporale etc.

Lumea gândită se scindează astfel în două niveluri care sunt totodată două funcționalități: a) *cosmosul noetic-fenomenal* constituit din ansamblul conceptelor necorespondente cu lumea reală (lume existentă, *posibilizată*); b) *cosmosul noetic-nefenomenal* constând din ansamblul conceptelor necorespondente cu lumea reală (lumea rațional posibilă, dar *neposibilizată*). Conceptele cu existență pur rațională (fără corespondent faptic identificat) pot fi plăsmuiri mentale având un caracter pregnant exprimabil. De exemplu:  $\sqrt{2}$ ,  $\pi$ ,  $i$ ,  $\infty$  etc.

Conceptul ideal „nu poate fi pierdere a realului, ci doar potențare a lui sau regăsire a lui pe un nivel superior“ (C. Noica). Datorită idealizării, omul poate avansa până la un „comportament simbolic“ care-l ajută să depășească condiția de prizonier al câmpului perceptiv, dobândind o libertate și o independență în raport cu datele experienței senzoriale.

Operațiile nespecifice ale inteligenței umane – analiza și sinteza, comparația etc. – transferate în planul idealizării, având atributele gândirii metafizice și metapsihice, devin *metaoperații*: meta-analiza realizează o despărțire prin înălțare și depășire, meta-comparația care detașează analogii și disanalogii dincolo de empiric, meta-sinteza este operația unificatoare de gradul doi care se poate metamorfoza în complementarizare ca sinteză creatoare a opuselor.

Toate operațiile și meta-operațiile la care ne-am referit întrețin și dezvoltă *idealizarea* care reprezintă cooperarea sinergică, adică „împreună și deodată“ și „una prin alta, nu una după alta“ a tuturor operațiilor mintale (I. Mânzat, 1991).

## 6. „COMPLEXUL DE INFERIORITATE“ AL PSIHOLOGIEI ACTUALE

Psihologia contemporană suferă de un „complex de inferioritate“ în fața fizicii și a celorlalte științe exacte și foarte vechi. Psihologia îndrăznește mai puțin decât fizica să recunoască faptul că se folosește de experimentul mental și de idealizare. Fizica teoretică utilizează frecvent experimente mentale, experimente ideale și intuiții creatoare și nu ezită să recunoască. Chiar se mândrește, considerându-le virtuți ale spiritului științei.

Știința modernă și postmodernă ne învață că îndărătul vizibilului simplu se află un invizibil complicat, că ordinea ascunsă, profundă este mai subtilă și mai complexă decât ordinea manifestă, superficială. În acest sens afirma Paul Valéry că în cunoașterea modernă „miraculosul și pozitivul au contractat o uimitoare alianță“. Acest contract nu este valabil doar în domeniul științelor fizice și matematice. Ceea ce este adevărat pentru aceste științe este posibil adevărat și pentru altele ca antropologia, psihologia, sociologia. Ceea ce funcționează în științele fizice funcționează probabil și în științele omului. Este foarte greu să ne dăm seama de acest *transfer depărtat*. Este din pricină că în gândirea științifică din domeniul umanului s-au refugiat toate prejudecățile și frustrările, inclusiv cele astăzi evacuate din științele exacte. Într-un domeniu atât de apropiat de ei și atât de mișcător, cercetătorii, pentru a vedea „limpede“, au încercat să reducă totul la un sistem. S. Freud explică totul, G. Allport descrie și explică totul. J.P. Guilford crede că a găsit structura intelectului (faimosul „cub“ al lui Guilford). Psihismul devine, sub impulsul psihologiei cibernetice, *sistem psihic uman* în care totul pare perfect și clar organizat. În fond, psihismul nu poate fi conceput ca sistem, întrucât în câmpul lui haosul este permanent și dinamic; putem accepta ideea că psihismul uman tinde spre sistemicitate sau că este un „sistem în devenire“ și incomplet. Acesta este cel mai malefic „complex de inferioritate“ al psihologiei actuale, care foarte ușor se transformă în „complex de superioritate“ (vezi psihologia cognitivă și psihologia transpersonală).



Psihologia se teme, mai mult decât alte științe, să nu fie acuzată iar de ocultism, ezoterism, misticism. Dar nu se sfiește să facă o religie din selfism, cognitivism sau transpersonalism.

## **7. EXPERIMENTUL MENTAL POATE FI O CALE DE ACCES ÎN ORIZONTUL METAPSIHIC ȘI PARANORMAL**

Prin complexe și multiple experimente mentale fizicienii contemporani au ajuns la o imagine-concept cunoscută sub numele de *paradigma vibratorie* (inspirată probabil din „Nada Brahma” indiană după care muzica, armoniile sunt telosul lumii). Această paradigmă a condus la ipoteza – de neacceptat pentru pozitiviștii încrezători doar în rezultatele și măsurătorile experimentului factual – că nu este deloc neștiințific a admite și o altă realitate a lumii decât cea percepută. Biochimistul Rupert Sheldrake (1985) a făcut senzație cu teoria sa a „câmpurilor morfogenetice” după care orice structură ordonată din cosmos exercită o influență, adică are un efect asupra restului cosmosului. Efectul constă în aceea că, după apariția structurii respective, multiplicarea, proliferarea ei devin mai probabile. Existența unei molecule deosebit de complexe, de pildă a unei enzime, înlesnește formarea altor asemenea molecule. Prin urmare: efectul până acum neglijat al oricărei structuri asupra întregii ambiante constituie un „câmp morfogenetic”. Plantele și animalele determină, prin simpla lor existență, o creștere a probabilității reapariției polifiletice a unor organisme de aceeași structură. Instinctul, învățarea se explică tot prin câmpuri morfogenetice. Dacă un animal asimilează o informație, altor exemplare din aceeași specie le vine mai ușor să-și însușească informația respectivă. Dar teoria lui Sheldrake se mai sprijină și pe o altă idee, una mai mult orientală, care se deosebește de concepția aristoteliană, fiind numită „cauzare supratemporală”, cauzare cumulativă. *Forma* imuabilă dobândește posibilitatea transformării. Câmpurile morfogenetice devin obiect al transformării și al proceselor de învățare. „Forma” nu

rămâne fixă, nici pur iterativă. Ea se schimbă, „căci tot ce se petrece realmente în lume va determina generarea morfologică transmisă prin câmpurile morfogenetice“. La Platon și Aristotel procesul este monodimensional: lumea reflectă formele date ca idei. La Sheldrake chiar și lumea formelor manifestate se reflectă în procesele morfogenetice, ceea ce înseamnă că are loc o intercondiționare și o influență reciprocă. Procesul este cumulativ, dincolo de timp și peste generații. O asemenea idee lipsește din tradiția spirituală occidentală, iar Sheldrake știind de unde s-a inspirat, o asemuiește cu principiul Karmei, cu budismul Mahayana, care trimite la un gen de *memorie cosmică*. Tot ceea ce se petrece în univers este un fel de Karma cosmică.

Mihai Drăgănescu (1990) demonstrează existența unui „inel al lumii materiale“ constituit dintr-un circuit de comunicare continuă între universul vizibil și „profunzimile materiei“, o structură a cărei intuiție deschide orizonturi nebănuite pentru cunoaștere și instituie o „legătură directă între fizic, viu și psihism“. David Bohm (1987) o numește „ordine implicată“ și o așează la baza lumii vizibile. Din ea derivă spațiul, timpul și particulele elementare. Dar mai există, spune el, și o „supraordine implicată“, mai subtilă, mai inefabilă, care generează prima ordine implicată și o ajută să se organizeze. Este *totalitatea totului* care organizează sensurile.

Dumitru Constantin (1981) arată că telepatia, hipnoza, câmpul energetic al organismului, sensibilitatea plantelor, structura subconștientului ș.a. primesc explicații naturale prin universul informațional al celulei vii (așa numita „percepție primară“). De aici pulsația ritmică a organismelor monocelulare, reacția senzitivă a plantelor la stimuli externi, autoreglarea organismului viu și alte fenomene „normale“, dar considerate până nu demult ca insolite. Analizând câteva experiențe liminale și subliminale, supravegheate științific, Dumitru Constantin ne îndeamnă la meditație, la reflecție asupra modului nostru de a gândi natura și a ne gândi pe noi înșine: suntem suficient de pregătiți psihic pentru a primi informații care șochează modul nostru obișnuit de gândire științifică, spre a scruta în profunzime cauzele unor fenomene inexplicabile rațional și chiar a ne elibera de obișnuința abordării acestora în contextul categoriilor de cauză, timp, spațiu?

Cercetările psihologice actuale asupra *stărilor de conștiință modificată* (B. Wolman și M. Ullman, 1986) se referă la modificarea calitativă a întregului model mental în raport cu cel cotidian. Percepția spațio-temporală dispare, omul își pierde autocontrolul, expresia emoțională se schimbă, imaginea corporală este alta, semnificațiile sunt răsturnate, se instalează sentimentul inefabilului și se schimbă întregul mod de gândire. La toate acestea cercetarea psihologică transpersonală mai adaugă: dispariția continuum-ului liniar dintre micro și macrocosm, transcenderea limitelor dintre substanță, energie și conștiință, experiențe complexe cu conținut condensat, transcenderea deosebirii dintre Self și lumea externă. Stările de conștiință schimbate pot fi provocate prin reducerea de stimuli și excitații sau prin suprastimulare senzorială, prin luciditate sporită, autohipnoză etc. Experimentul pozitiv nu mai este suficient în aceste cazuri. Intervin forme subtile de experimentare mentală sau experimente mixte, complementare pozitiv-mentale.

Simbologul Vasile Lovinescu (1994, p. 75) aduce un elogiu spiritualității umane: „Este legitim să aștepți în fiecare clipă un miracol, pentru că însăși clipa este un miracol. Sub aparența că aștepți un miracol, vrei în realitate să reintri în normal... Totul e să nu aștepți miracolul din afară; trebuie să-l pândești și să-l soliciți în tine însuși, fiindcă, chiar atunci când vine din afară, în realitate exteriorizează o minune lăuntrică“.

Știința postmodernă și știința viitorului va fi o știință care mai mult hominizează decât materializează. P. Teilhard de Chardin crede: „Adevărata fizică este aceea care va izbuti să integreze *Omul total* într-o reprezentare coerentă a lumii“. Astăzi cercetătorii experimentalisti încep să înțeleagă că un studiu aprofundat și „cu alți ochi“ (Carl Sagan) al cărții naturii cere și altceva decât spirit de observație, spirit științific, experiment fizic, măsurători exacte, evaluări matematice și chiar altceva decât inteligență. Ar trebui ca în stadiul în care se află știința actuală spiritul uman să se întreacă pe sine, ca inteligența umană să se depășească. Omenesc – prea omenescul nu mai ajunge. Modernul întârziat este raționalist. Contemporanul viitorului se simte transcendent și cosmizat.

- Henri Bergson, *Energia spirituală*, București, Edit. Antet, 1994.
- Claude Bernard, *Introducere în studiul medicinei experimentale*, București, Edit. Științifică, 1958.
- Lucian Blaga, *Cenzura transcendentă* („Trilogia cunoașterii“, III), București, Edit. Humanitas, 1993.
- David Bohm, *Wholeness and the implicate order*, London, Routledge and Kegan Paul Ltd. 1987.
- David Bohm and Jean Pierre Vigier, „Model of the causal interpretation of quantum theory in terms of a fluid with irregular fluctuations“, *Physical Review*, 1954, vol. 96.
- Fritjof Capra, *The Tao of Physics*, New York, Edit. Bentam, 1983.
- Sorin Comoroșan, *Exerciții de naivitate*, București, Edit. Cartea Românească.
- Dumitru Constantin, *Inteligența materiei*, București, Edit. Militară, 1981.
- Mihai Drăgănescu, *Informația materiei*, București, Edit. Academiei Române, 1990.
- George Enescu, *Filozofie și logică*, București, Edit. Științifică, 1973.
- Sigmund Freud, *Viața mea și psihanaliza*, Iași, Edit. Moldova, 1993.
- Nae Ionescu, *Curs de logică*, București, Edit. Humanitas, 1993.
- Vasile Lovinescu, *Jurnal alchimic*, Iași, Institutul European, 1994.
- Pierre Malrieu, *La construction de l'imaginaire*, Bruxelles, Edit. Ch. Dessart, 1967.
- Ion Mânzat, „Funcționalități specifice ale transferului din învățarea structural-construcțivă“, *Revista de psihologie*, 1988, nr. 3.
- Ion Mânzat, „Deschiderea sinergeticii. Spre un orizont teoretic neclasic în psihologie“, *Revista de psihologie*, 1991, nr. 1-2.
- Ion Mânzat, „Pluralitate și postmodernism în psihologie“, *Revista de filosofie*, 1994, nr. 4.
- Ion Mânzat, *Intuiția ca psihosinerergie*, comunicare la Sesiuneade comunicări științifice a Institutului Național de Informații, București, 17-18 martie 1995.
- Ion Mânzat, *Transconștiința umană. Introducere în psihologia credinței religioase*, București, Edit. Pan Publishing House, 1995 (sub tipar).
- Constantin Noica, *Jurnal de idei*, București, Edit. Humanitas, 1993.
- Sidney Parnes, *Creative Behavior Guidebook*, New York, Ch. Scribner's Sons, 1967.
- Jean Piaget, *Psihologia inteligenței*, București, Edit. Științifică, 1965.
- Paul Popescu-Neveanu, *Curs de psihologie generală*, vol. II, București, Edit. Universității București, 1977.
- Paul Popescu-Neveanu, *Dicționar de psihologie*, București, Edit. Albatros, 1978.
- Jean Paul Sartre, *L'imagination*, Paris, Ed. Gallimard, 1940.
- Rupert Sheldrake, *Das schöpferische Universum. Die Theorie des morphogenetischen Feldes*. München, W. Goldman Verlag, 1985.
- Liviu Sofonea și N. Ionescu-Pallas, *Gândirea blagiană despre „suprametodă“ și unele particularități metateoretice ale fizicii moderne*, în :Dumitru Ghișe, Angela Botez, Victor Botez (coord.), *Lucian Blaga – cunoaștere și creație* (culegere de studii), București, Edit. Cartea Românească, 1987, p. 254-276.
- Roger W. Sperry, „Some effects of disconnecting the cerebral hemispheres“ *Science*, 1982, p. 217-.
- Benjamin Wolman and Montague Ullman, *Handbook of states of consciousness* New York, Van Nostrand Reinhold Co., 1986.

# COERENȚA MENTALULUI ȘI FORMALULUI ÎN EXPERIMENT

Gheorghe ȘTEFAN

(Universitatea Politehnica București)

Într-o lume în care experimentul practicat sistematic se extinde dincolo de limite tradiționale, devenind **experiment mental** (într-o manieră consistentă prin *experiment filosofic* [Drăgănescu '79]) sau **experiment formal** (ilustrat semnificativ prin *experimentul matematic* [Chaitin '94]), ne putem pune întrebarea dacă mai există un domeniu în care experimentul să nu fie o componentă esențială a comportamentului uman. Experimentul este un reflex pe care omul nu l-a evitat niciodată, dar nici nu l-a conștientizat întotdeauna. Atunci când *mintea* omului și *formele* pe care ea le imaginează, selectează sau construiește, intră sub incidența unui joc esențial pot apare tensiuni date de neconcordanțe ce se cer explicate. **Complexitatea** devine astfel un concept cheie pentru a limpezi relația dintre deschiderile mentalului și închiderile formalului. Ipoteza pe care o avansăm este aceea că există o necorelare între complexitatea proceselor mentale și complexitatea reprezentărilor formale. Se pare că va trebui să tindem către un spațiu al cunoașterii în care formele se pot deschide către neforme iar mentalul acceptă închideri ce obturează căi sterile.

## 1. CĂTRE UN NOU FORMAL

Relația dintre formal și neformal este intens discutată în epistemologia contemporană, în primul rând datorită noilor aspecte evidențiate de studiul calculabilității și de dezvoltarea tehnolo-

giilor informatice. Atunci când rezultatele unui domeniu devin spectaculoase – așa cum s-a întâmplat, spre exemplu, cu știința calculatoarelor – se ridică, din cele mai diverse motive, întrebări asupra fundamentelor acestor rezultate. Omul pare a se teme de performanța ce depășește anumite limite. Nu de puține ori aceste temeri sunt fundamentale. O atitudine lucidă va clasifica eventualele limite astfel încât să se poată crea baza convertirii unora în deschideri către noi dezvoltări utile.

## 1.1. LIMITE BENIGNE

Luminile și umbrele abordării formale au fost dintotdeauna în atenția celor ce studiau fundamentele științei. Spre deosebire de aceștia, cei ce aplicau știința fundamentată formal, mai apropiați de cenușiul realității, s-au manifestat mai puțin exaltat sau pesimist. Succesul diverselor formalisme în rezolvarea unor probleme concrete sugerează faptul că limitele formalizabilului au un caracter benign, ele pot fi tolerate atunci când concretețea soluțiilor poate servi drept cheazășie.

Din acest motiv *teorema de incompletitudine* a lui Kurt Gödel, chiar dacă se manifestă ca o limită fundamentală a sistemelor formale ce depășesc o anumită complexitate, are un caracter benign, în sensul că implicațiile ei pot fi ocolite în utilizarea sistemelor formale. Chiar atunci când se manifestă sub forma unor probleme insolubile în știința calculatoarelor ea nu înspăimântă nici un informatician. Într-adevăr, *problema opririi*, ce planează asupra fiecărui cuplu *program – date de intrare*, nu a putut bloca spectaculoasa dezvoltare a informaticii. Construcții fabuloase cum ar fi *big omega* [Chaitin '78], un număr a cărui reprezentare binară conține biți independenți, ne surprind intelectual dar nu ne împiedică să folosim cu succes construcțiile formale.

Să nu uităm sintagma clasică: *eu mint*. Faptul că nu înțelegem la ce se referă, atunci când este rostită de un individ, nu ne împiedică să folosim în continuare limbajul natural.

Ne-am acomodat, prin demersuri ce provin din cele mai diverse spații, cu ideea că adevărul se extinde și dincolo de ceea ce este demonstrabil. Se pare, chiar, că demonstrabilul este un subdomeniu cu o extindere cantitativă nesemnificativă în raport cu adevărul. Limitrea la demonstrabil, chiar dacă unora le pare opresivă, este acceptată de o atitudine pragmatică. Dar, chiar o atitudine metafizică cred că este reconfortată la gândul că există adevăruri dincolo de uscăciunea demonstrabilului. Amploarea și profunzimea acestor adevăruri de dincolo de demonstrabil transformă limita într-o deschidere.

De la paradoxul mincinosului și până la problema opririi istoria celor create de mintea omului nu a făcut decât să dea semne discrete, prin marcarea unor limite benigne, despre o realitate de dincolo de formal în care adevărul se manifestă nestingherit de rigoarea atrofiantă a formalismelor. Mintea poate experimenta deci într-un spațiu al adevărului mult extins peste cel al demonstrabilului, generându-se o tensiune datorită faptului că formalul trebuie să se limiteze la demonstrabil.

Deci, o primă țintă strategică pe care ne-o putem propune este aceea de a extinde cantitativ spațiul demonstrabilului pentru a micșora incompatibilitatea cu amploarea spațiului adevărului în care mentalul se mișcă cu dezinvoltură.

## 1.2. LIMITE MALIGNE

Spre deosebire de limitele fundamentale, ce se pot converti în deschideri, limitele cantitative puse de rezolvarea unor probleme concrete nu pot fi valorificate într-o manieră la fel de pozitivă. Atunci când demonstrăm că adevărul unei construcții mentale nu este decidabil ne lovim de un perete dincolo de care *sperăm* că se află ceva. Dar atunci *știm* cum se rezolvă o problemă, importantă pentru optimizarea existenței noastre, dar constatăm că *nu avem suficiente resurse* pentru a ajunge la soluție, ne aflăm în situația de neînviat a celui care străbate o peșteră al cărei tavan există dar nu se întrevește.

Un prim exemplu este *problema comisvoiajorului*. Un comisvoiajor trebuie să viziteze  $n$  orașe, cunoaște distanțele dintre ele și trebuie să-și aleagă un circuit minim care trece o singură dată prin toate orașele. Soluția teoretică, sub forma unui program, este foarte ușor de găsit, dar se demonstrează că pentru cele mai bune calculatoare existente, dacă  $n > 64$  (o cifră reală rezonabilă) soluția optimă nu poate fi calculată decât în multe mii de secole.

Un al doilea exemplu din aceeași categorie îl constituie *problema satisfiabilității*: dându-se o funcție booleană  $F(x_1, \dots, x_n)$ , de  $n$  variabile, să se afle dacă există cel puțin o configurație binară de intrare pentru care funcția  $F$  ia valoarea de adevărat. Numărul de configurații binare de intrare este  $2^n$ . Programul ce caută soluția problemei este foarte simplu dar nu poate evita calcularea funcției  $F$  pentru toate configurațiile binare de intrare. Deci timpul de rulare este proporțional cu  $2^n$ . Pentru  $n = 64$  și pentru un calculator capabil să calculeze valoarea funcției  $F$  într-o milionime de secundă (o performanță notabilă!) putem să așteptăm aproximativ 5 000 de secole pentru aflarea rezultatului.

Stephen Cook a demonstrat că există o clasă foarte largă de probleme concrete echivalente ca efort de calcul cu problema satisfiabilității. Este vorba de clasa problemelor NP-complete (a acelor probleme care nu se pot rezolva pe o mașină finită decât într-un timp proporțional cu  $2^n$ , unde  $n$  este proporțional cu dimensiunea datelor de intrare). Rezolvarea oricăreia cu un algoritm mai simplu, de tip polinomial (un algoritm este polinomial dacă rezolvă problema într-un timp proporțional cu  $n^c$ , unde  $n$  este proporțional cu dimensiunea datelor de intrare iar  $c$  este o constantă) ar atrage după sine rezolvarea simplificată a tuturor problemelor NP-complete.

Să nu poți atunci când știi este mult mai frustant decât să nu știi pentru că nu se poate. Cu toate că nu suntem în fața unor limite teoretice fundamentale, putem afirma caracterul malign al acestor imposibilități care nu deschid către nimic altceva.

Este stupefiantă simplitatea enunțului problemei satisfiabilității. Cu atât mai blocantă este imposibilitatea de a găsi o soluție eficientă și obligativitatea de a accepta soluția banală a căutării soluției prin simpla încercare a tuturor combinațiilor de intrare posibile. Pentru atâtea alte probleme s-au găsit algoritmi sofisticati care



optimizau găsirea soluției și pentru această problemă aparent banală acest lucru nu este posibil? „Simplitatea spațiului” în care a fost definită problema credem că este, măcar parțial la originea dificultăților. Principiile calculului boolean sunt foarte simple: ele pot fi expuse complet pe una sau două pagini. Această simplitate stă la baza construcțiilor fizice concrete (circuitele logice) ce permit structurări ample, materializate sub forma calculatoarelor moderne. Această consecință strict pozitivă nu este cumva acompaniată, „într-un târziu”, de imposibilitatea de a rezolva concret o problemă de genul problemei satisfiabilității? Simplitatea inițială a reprezentărilor se „răzbună” prin complexități inacceptabile ale unor consecințe importante.

Vom încerca un răspuns indirect, făcând unele evaluări cantitative într-un spațiu definit cu o simplitate comparabilă. Să considerăm lumea imaginilor alb-negru ce pot fi desenate pe  $1 \text{ cm}^2$ , cu o rezoluție de 32 de linii, fiecare având 32 de puncte. Este vorba de lumea trunchiată a imaginilor ce pot fi desenate pe  $1 \text{ cm}^2$  cu o peniță de aproximativ 0,3 cm grosime, folosind numai linii verticale sau orizontale și puncte. Deci, oricare din aceste imagini poate fi prezentată prin  $32 \times 32 = 1024$  de semne binare (0 pentru alb și 1 pentru negru). Rezultă că numărul de imagini *potențiale* distincte ce pot fi astfel reprezentate este de  $2^{1024} \approx 10^{320}$ . Să calculăm acum câte imagini din cele astfel posibile ar putea să fie actualizate la o scară de ordinul de mărime al civilizației umane, pentru a vedea în ce măsură o astfel de reprezentare este utilă. Considerăm că un om poate distinge maximum 10 imagini diferite pe secundă. Vom considera, mult acoperitor, că o populație de 10 miliarde de locuitori privește continuu, zi și noapte, imagini distincte, timp de 4 miliarde de ani. Numărul de imagini distincte ce s-ar putea stfel *actualiza* este:

$$\begin{aligned}
 &10^{10} \text{ (persoane)} \times \\
 &10 \text{ (imagini discernabile pe secundă)} \times \\
 &3\,600 \text{ (secundă pe oră)} \times \\
 &24 \text{ (ore pe zi)} \times \\
 &365 \text{ (zile pe an)} \times \\
 &4 \times 10^9 \text{ (ani)} \approx 10^{28} \text{ imagini.}
 \end{aligned}$$

Rezultă o concluzie stupefiantă: potențialul reprezentărilor se actualizează într-o proporție de aproximativ  $10^{-292}$ , adică într-o *proporție practic nulă*. Dar mașini pentru reprezentarea acestor imagini și teorii pentru tratarea lor se fac ca și cum toate aceste imagini ar fi posibile. Disproporția zdrobitoare dintre posibil și realizabil este însă prea mare pentru a o neglija. Teoriile pe care le concepem pentru a studia formal aceste imagini și mașinile pe care le construim pentru a le manipula sunt destinate tuturor celor aproximativ  $10^{320}$  imagini ignorând că, de fapt, cele ce vor fi efectiv actualizate nu sunt decât  $10^{28}$ . Reprezintă această infimă parte o clasă cu anumite caracteristici distincte în mulțimea tuturor imaginilor posibile? Dacă da, atunci reprezentarea ar trebui să pornească de la ele, cu consecințe ce ar putea fi foarte importante pentru „teoria imaginii de 1024 de pixeli”.

Am putea să ne justificăm reprezentarea binară pe care acum o folosim prin simplitatea teoriei și a mașinilor asociate. Dar nu ne putem reprimă întrebarea dacă nu cumva o altă reprezentare, eventual mai complexă, nu ar putea să fie mai adecvată și mai eficientă în cazul unor probleme realmente complexe? Bitul și funcțiile booleene sunt oare cele mai eficiente reprezentări pentru toate problemele pe care acum le soluționăm prin algoritmi cu complexitate exponențială? Există o șansă ca răspunsul să fie negativ.

### 1.3. SIMPLIFICAREA TEORIEI PRIN COMPLEXIFICAREA PRINCIPIILOR

De ce teoriile formale pot da seamă numai de adevăruri atât de simple în raport cu adevărurile pe care mintea le poate manipula? Credem că răspunsul este foarte simplu: omul de știință din toate timpurile a fost obsedat de două lucruri:

– *fundamentarea oricărui demers formal pe un număr cât mai mic de principii cât mai simplu formulate,*

cu speranța că pornind de la ele

– *se pot elabora construcții formale oricât de complexe.*

Dacă prima aspirație a fost cu prisosință realizată în foarte multe cazuri, cea de a doua nici măcar nu s-a dovedit justificată. Dezvoltarea *teoriei algoritmice a informației* a permis demonstrarea unei teoreme foarte importante [Chaitin '77], conform căreia:

*cu aproximația unei constante finite, nici o teoremă nu poate conține mai multă informație decât cea conținută în setul de axiome ce fundamentează teoria ce a permis deducerea ei.*

Deci, pornind de la teorii simplu și compact definite nu se pot deduce teoreme (adevăruri) care să depășească complexitatea teoriei cu mai mult de o constantă. Aspirația către simplitate a teoriilor clasice se dovedește o limită cantitativă esențială pentru realitățile de care ar putea da seamă. Cu o teorie formală simplu fundamentată nu putem să imaginăm decât experimente mentale limitate la adevărurile de o simplitate corespunzătoare pe care aceasta le poate manipula. Minte va trebui să se limiteze la demonstrabilitatea limită a unei teorii simple. Aspirația, cu valențe estetice, pe care au avut-o oamenii de știință din toate timpurile pentru evidențierea unor principii simple, ce se dovedesc utile și clare, a avut drept consecință tensiunea pe care astăzi o constatăm între mental și formal în experiment. Aceste teorii se dovedesc a fi inutile și neclare pentru rezolvarea unor probleme ce au devenit curente pentru omul de știință contemporan [Ștefan '95].

Spre exemplu, putem să interpretăm comportamentul unui individ într-un context dat, pornind de la modul în care funcționează sistemul său nervos, ținând cont de totalitatea proceselor fizice și chimice implicate. Dimensiunea travaliului, perfect posibil, întreprins pe această cale ar fi atât de mare și rezultatul atât de aproximativ încât efortul nu va putea fi justificat. Mult mai utilă ar fi evidențierea unor scenarii posibile semnificative pentru o clasă suficient de largă de comportamente, prin utilizarea cărora să poată fi oferită o interpretare, cu bună aproximație corectă, pentru comportamente umane concrete. O astfel de abordare, promovată de Radu Bogdan, permite soluții cu o complexitate rezonabilă pentru situații reale foarte nuanțate. Premizele unei astfel de teorii sunt evident mult mai ample decât ale celei ce pornește de la procese fizico-chimice elementare, dar

posibilitatea de a soluționa probleme reale, cu un efort rezonabil, crește foarte mult. *În cadrul unei teorii mai complex fundamentate construcțiile pot fi mai simple.*

Un alt exemplu îl pot constitui *logicile modale*, dezvoltate în secolul XX, prin intermediul cărora realități bine circumscrise pot fi mult mai simplu descrise decât folosind logica aristotelică.

Dezvoltarea *lingvisticii computaționale* se constituie într-un alt exemplu semnificativ. Un eșantion finit, dar semnificativ al limbii engleze – *corpus-ul* acestei limbi – a fost utilizat pentru evidențierea unui număr de aproximativ 30 de mii de reguli. Aceste reguli descriu, cu o acuratețe suficient de mare orice text scris în limba engleză, constituindu-se în „teoria formală a limbii engleze”. Această teorie nu mai respectă aspirația către simplitate, anterior menționată, dar permite analizarea sau construirea unor texte de mare complexitate scrise în limba engleză. Evident, această „teorie” nu poate fi folosită de un utilizator uman, ea este destinată exclusiv unei mașini puternice de calcul. Zeci de mii de reguli, corespunzând axiomelor și teoremelor unei teorii clasice, nu pot fi manipulate formal de un utilizator uman. Dar comportamentul calculatorului ce utilizează această teorie exersând un veritabil experiment, are o complexitate care poate deveni comparabilă cu cea a unui experiment mental realizat în același domeniu. Chiar dacă mintea mai posedă și conexiuni pe care un proces formal gândit și instituit nu le poate avea [Drăgănescu '79], distanța dintre un experiment mental și unul formal, realizate în contextul oferit de lingvistica computațională, sau abordări similare, se reduce semnificativ datorită creșterii spectaculoase a complexității experimentului formal.

Se pare că cheia către acomodarea mentalului cu formalul este renunțarea la carteziana aspirație către simplitate și claritate. **Utilitatea unei teorii depinde, în epoca informatizării cunoașterii, de complexitatea ei și nu de simplitatea care permitea manipularea comodă a unor principii simple.** Vom construi din ce în ce mai mult teorii pentru uzul mașinilor. Implicarea mentalului va presupune într-o măsură din ce în ce mai mare imaginarul, a cărui exersare a fost în mod regretabil neglijată în cultura europeană odată cu reforma protestantă [Culianu, 1994]. Raționalismul cartezian a încercat mecanizarea gândirii și a reușit crearea unor

mecanisme pentru exersarea performanță a gândirii formale: calculatoarele. Astfel omul se regăsește liber să-și exerseze imaginarul la care a trebuit să renunțe parțial timp de câteva secole.

## 2. PRINCIPALA PROVOCARE: COMPLEXITATEA

Redobândirea accesului nestingherit la imaginar, prin care sper să putem reechilibra relația dintre mental și formal, este încă un proces în curs de desfășurare. Motorul principal al acestui proces este provocarea oferită de **complexitate**. Drumul către imperiul imaginarului este lung, dar îl parcurgem din ce în ce mai rapid, fugăriți de un monstru, cu atât mai fioros cu cât este mai greu de descris: complexitatea.

### 2.1. COMPLEXITATEA: UN CONCEPT CE NU SE LASĂ UȘOR DEFINIT

Nu de puține ori cunoașterea a evoluat sub presiunea *sugestiei* oferită de „concepte” imprecis definite. Sugestia vagului poate fi uneori imperativă, într-o situație paradoxală, generatoare de imprevizibile cristalizări. Acesta poate fi motivul pentru care cunoașterea, într-o manieră „inconștientă”, tolerează sugestia imperativă a unor concepte ce scapă definițiilor unanim acceptate. **Timpul și informația** intră în categoria acestor concepte catalizante. Un altul tinde să capete un statut similar: **complexitatea**.

Am operat intens cu acest concept chiar în textul de față, mizând pe sugestia lui pentru unele din cele mai importante idei pe care am încercat să le impun. Întregul demers poate fi pus sub semnul întrebării dacă ne vom încrâncena în adoptarea unei definiții riguroase pentru ideea de complexitate. Literatura de specialitate ne oferă mai multe căi, vom propune și noi una, dar

rămânem la gândul că o definiție riguroasă unanim acceptată nu există încă și poate ar fi chiar inutilă. Calea către o definiție care nu există este de multe ori o modalitate superioară de cunoaștere ce nu acceptă încremenirea atrofiată în formă.

Convergența „asimptotică” spre formă poate fi rezultatul unei negocieri subtile pe care mentalul o face cu formalul pentru a putea participa consistent la un experiment comun.

## 2.2. DEFINIȚII ALE COMPLEXITĂȚII

În funcție de domeniul de aplicație sau de profunzimea abordării, conceptul de complexitate a fost definit în diferite moduri. Vom lista o parte dintre acestea [Edmonds '95] pentru a pregăti introducerea definiției noastre în paragraful următor.

**Complexitatea ca mărime** (size) este uneori folosită atunci când interconexiunile domeniului evaluat nu sunt importante sau când evaluările ce se fac sunt grosiere. De multe ori se folosește termenul de *complexitate a calculului* pentru a desemna aspecte cantitative ușor de evidențiat, cum ar fi: timpul de calcul (în cicluri mașină) sau dimensiunea memoriei folosite. Sistemele digitale sunt uneori evaluate prin complexitatea văzută ca număr de intrări în toate circuitele ce compun sistemul, ignorându-se aspecte legate de topologia sau funcția circuitului. Mărimea substituită complexității crează de multe ori confuzii ce nu pot fi evitate decât prin definiții mai nuanțate.

**Complexitatea ca mărime a definiției minime**, numită și **complexitate algoritmică**, oferă distincții mai nuanțate. Teoria algoritmică a informației promovată în anii '60 de Kolmogorov, Solomonov și Chaitin, dezvoltată amplu de cel din urmă [Chaitin '87], permite și evaluări cantitative ale unor realități compuse din multe elemente. Exemplul cel mai simplu, la care se pot reduce toate evaluările formale, este cel a unui șir binar, de regulă, foarte lung. Un astfel de șir este *simplu* dacă, spre exemplu, este format de repetarea de 500 000 ori a secvenței binare 0011. Un șir este *algoritmically complex* dacă este format din 2 milioane de biți

înlănțuiți aleator, fără nici o regulă. În primul caz descrierea cea mai simplă a șirului este: *generează de 500 000 ori 0011*, iar în cel de al doilea caz descrierea cea mai simplă este: *generează 0100011100 ... 110010100001101*, unde șirul binar conține 2 milioane de semne. Mărimea primei definiții este de o jumătate de rând, pe când cea de a doua definiție umple spațiul a câtorva volume. Conform acestei definiții cel de al doilea are complexitate maximă, neavând o reprezentare mai compactată. Rămâne să ne punem întrebarea dacă: cel de al doilea șir reprezintă într-adevăr ceva? Rămâne totuși important faptul că se poate face o estimare cantitativă corelată și cu conținutul reprezentării.

Este interesant de amintit și punctul de vedere, preponderent calitativ, al lui B. Edmonds, conform căruia complexitatea este: *That property of a language expression wich makes it difficult to formulate its overall behaviour, even when given almost complete information about its atomic components and their inter-relations.*

### 2.3. O DEFINIȚIE SEMANTICĂ A COMPLEXITĂȚII

Principala critică pe care o facem definițiilor anterioare este aceea că nu au ținut cont în definirea complexității de aspectele *semantice*. Teza pe care o avansăm este aceea că nu putem defini complexitatea unei reprezentări independent de *semnificația* pe care i-o acordăm. O abordare de acest tip ne va permite să ținem cont, în egală măsură de:

- *tendința către rigoare a formalului, prin căutarea unei măsuri*
- *apetența mentalului de a manipula înțelesuri*

în tentativa de a concilia formalul și mentalul, în disputa pentru a „impune” o definiție care să poată fi utilă în experimente proprii. O definiție pur cantitativă este inutilă pentru că uniformizează aspecte calitative pe care mentalul se sprijină în demersul său. De asemenea, o abordare pur calitativă nu poate oferi abordării formale nici un reper. Avem nevoie de un concept care să asigure un context acceptat în egală măsură de lumea mentalului ca și de cea a formalului.

Definiția pe care o propunem necesită unele precizări preliminare.

1. Orice *reprezentare* poate fi adusă sub forma unui *șir de simboluri binare*.

2. Mărimea unei reprezentări este dată de lungimea șirului binar asociat.

3. Diferența a două reprezentări est definită numai pentru reprezentări de dimensiune egală și are drept rezultat funcția *sau-exclusiv* realizată între biții ce ocupă poziții identice. (Exemplu: diferența dintre:

0001110010100011

și

1100110011001100

este:

1101000001101111.

După cum se observă s-au modificat (complementat) acei biți „descăzutului“ care au corespuns l-urilor „scăzătorului“.)

Putem propune următoarea definiție:

**Complexitatea semantică** a unei reprezentări est dată de modulul diferenței dintre complexitatea algoritmică a acelei reprezentări și complexitatea algoritmică maximă pe care o poate avea o reprezentare care poate fi scăzută din reprezentarea inițială fără ca semnificația acesteia să poată fi modificată.

Vom încerca să explicăm această definiție folosind ca exemplu patru reprezentări bidimensionale:

- (a) o tablă de șah
- (b) desenul unui circuit electronic format din zeci de componente electronice (rezistențe, condensatoare și tranzistoare)
- (c) un desen de Picasso
- (d) un număr aproape egal de puncte albe și negre distribuite aleator.

**Complexitatea algoritmică** ( $C_A$ ) a acestor reprezentări este, în aceeași ordine:

- (a) foarte mică
- (b & c) moderată
- (d) foarte mare.



Semnificația primei reprezentări (tabla de șah) se poate păstra dacă un număr mic de biți se completează. Desenul circuitului electronic își păstrează semnificația dacă din el se scade o imagine relativ simplă. Desenul de Picasso nu suportă nici o modificare dacă vrem ca semnificația lui – „desen de Picasso” – să nu fie alterată. Ultima reprezentare rămâne la fel de „semnifictivă” – adică nesemnificativă – chiar dacă i se completează toți biții.

În consecință, *complexitatea semantică* ( $C_S$ ) a celor patru reprezentări va fi:

- (a)  $C_S \approx C_A$ , pentru că din  $C_A$  se scade o mărime comparabilă
- (b)  $C_S \approx C_A$ , pentru că din  $C_A$  se scade o mărime nesemnificativ de mică
- (c)  $C_S = C_A$ , pentru că din  $C_A$  nu se poate scădea decât o reprezentare nulă
- (d)  $C_S = 0$ , pentru că din  $C_A$  se poate scădea o valoare identică.

Complexitatea semantică poate fi mare pentru imagini ce au o reprezentare neregulată dar posedă o semnificație. Spre deosebire de complexitatea algoritmică unde o reprezentare aleatoare posedă o complexitate maximă, în cazul complexității semantice o reprezentare aleatoare neavând semnificație, complexitatea ei este nulă (deoarece rămâne la fel de nesemnificativă prin complementarea tuturor biților reprezentării). De asemenea, imaginile unice, de regulă cele cu valoare artistică, au complexitatea algoritmică egală cu cea semantică, deoarece nici o parte a reprezentării nu poate fi modificată fără a se modifica semnificația. Am putea spune că atunci când  $C_S = C_A$ , reprezentarea are pe lângă semnificație și *sens*. O astfel de reprezentare nu admite încadrarea în categoria formelor, putem spune despre ea că este *neformală*.

Am putea stabili următoarele corespondențe folosind exemplul dat prin cele patru figuri:

- figurile din categoria (a) sunt reprezentări *formale*
- figurile din categoria (b) sunt reprezentări *formalizabile*
- figurile din categoria (c) sunt reprezentări *neformale*
- figurile din categoria (d) nu sunt reprezentări.

Tranziția gradată între formal și neformal se poate face numai într-un spațiu în care semnificațiile arbitrează, printr-o experiență mentală, un proces imposibil de imaginat ca o experiență pur formală.

Definiția complexității sematice se obține într-un experiment în egală măsură mental și formal. De aici va rezulta și utilitatea acestui mod de a înțelege complexitatea pentru coerența mentalului și formalului în experiment.

### 3. COERENȚA MENTAL-FORMAL PRIN CORELAREA COMPLEXITĂȚILOR

Complexitatea semantică va permite relaxarea tensiunii dintre experimentul mental și cel formal, datorită compatibilității pe care o asigură evaluarea realizată cu o măsură „acceptată” de ambele părți. Reglarea diferenței, acum uriașe, dintre adevăr și demonstrabil va putea fi făcută cu un instrument de măsură comun.

Concluzia cea mai importantă ce se poate desprinde din acest text este aceea că *elementele unei teorii trebuie să fie suficient de elaborate pentru a avea din start o semnificație suficient de mare*. Dezvoltările teoretice clasice pornesc de la elemente și principii prea simple, care capătă semnificație numai pe măsură ce teoria este treptat elaborată. (Un exemplu limită îl constituie lumea imaginilor de  $1\text{ cm}^2$ , în care bitul nu are nici un fel de semnificație, motiv pentru care spațiul imaginilor posibile devine, din start și pentru totdeauna, absurd de mare. Dacă am proceda la fel în cazul lingvisticii, considerând ca structuri elementare literele alfabetului, cu care acceptăm formarea oricărei combinații posibile am ajunge la o aberație similară; nici măcar orice combinație de cuvinte nu este acceptată, fapt pentru care mulțimea textelor posibile este drastic restricționată.)

Tranziția de la jocul formelor nesemnificante către operarea cu forme din ce în ce mai încărcate de semnificații apropie experimentul formal de cel mental. Complexitatea semantică în cele două tipuri de experimente începe să convergă către valori apropiate. Construcții, adevăruri din ce în ce mai complexe intrând sub incidența abordărilor formale, tensiunea dintre mental și formal tinde să se atenueze.

Aroganța cu care teoreticienii încearcă să fundamenteze cât mai simplu teoriile pe care apoi doresc să-și sprijine experimentele mentale, formale sau reale, a condus la blocajul ce caracterizează relația dintre aceste trei tipuri de experimente. Deblocarea nu este posibilă decât prin *teorii fundamentate complex* („neelegant“) în care să se poată oferi *soluții simple* problemelor care astăzi beneficiază de soluții inacceptabil de mari consumatoare de timp și resurse fizice.

Tehnologiile informaționale sprijină această inversare radicală, anticarteziană, de viziune. Experimentul formal este eficient posibil numai pe suportul oferit de instrumentele informatice. *Mintea omului și instrumentele de calcul pot și merită să interacționeze numai dacă se implică specific în complexități apropiate.* Forța subtilă a imaginarului se poate actualiza numai sprijinită de forța brută a formalului. Totul este ca aceste forțe să poată fi ținute aproape una de cealaltă. Cine le îndepărtează? Cei ce cred că, dacă demersul cartezian a permis construcția calculatoarelor, atunci, același demers trebuie să caracterizeze și folosirea lor performantă.

## BIBLIOGRAFIE

- [Chaitin '77] Gregory Chaitin: „Algorithmic Information Theory“, în *IBM J. Res. Develop.*, Iulie, 1977.
- [Chaitin '87] Gregory Chaitin: *Algorithmic Information Theory*, Cambridge Univ. Press, 1987.
- [Chaitin '94] Gregory Chaitin: „The Limits of Mathematics“, preprint, IBM Research Division, 1994.
- [Culianu '94] Ioan Petru Culianu: *Eros și magie în Renaștere. 1984*, Editura Nemira, 1994.
- [Drăgănescu '79] Mihai Drăgănescu: *Profunzimile lumii materiale*, Editura Politică, București, 1979.
- [Edmonds '95] Bruce Edmonds: „What is Complexity? – The philosophy of complexity *per se* with application to some examples in evolution“, preprint, Centre for Policy Modelling, Manchester Metropolitan University, Manchester, UK; E-mail: b.edmonds@mmu.ac.uk.
- [Ștefan '95] Gh. Ștefan „Informatica și alternativa civilizație-cultură“, în *Revista română de informatică și automatică*, Vol. 5, Nr. 1, 1995.

# MENTAL ȘI EXPERIMENTAL ÎN ȘTIINȚELE COGNITIVE

G.G. COSTANDACHE

(Universitatea Politehnica, București)

„La métaphore permet de comprendre quelque chose et d'en faire l'expérience en terme de quelque chose d'autre“  
(G. Lakoff et M. Johnson)

## 1. ȘTIINȚELE COGNIȚIEI SAU INTELIGENȚA ARTIFICIALĂ?

Numeroși teoreticieni ai inteligenței artificiale sunt de acord să recunoască despre proiectele lor că nu urmăresc o simulare veritabilă a inteligenței umane, ci doar o „emulație“ (cum se exprimă Domenico Parisi, 1989), adică propun actualizarea unui comportament pur și simplu analogic. Dacă ar fi important numai aspectul pragmatic, respectiv, dacă ne-ar interesa numai rezultatele sau capacitatea artefactelor de a soluționa problemele pentru care au fost construite, acestea și-ar merita denumirea de *inteligente*, tocmai în măsura realizării performanțelor scontate. Deși soluțiile actuale nu sunt lipsite de unele dificultăți, departe de a reprezenta obstacole în dezvoltarea științelor cogniției sau descurajări pentru ambițiile teoreticienilor inteligenței artificiale, care urmăresc nemijlocit nivelul performanțelor, aceste dificultăți acționează ca *stimulent euristic* (J. –M. Besnier, 1993). În continuare, pornind de la astfel de considerații, vom încerca să aducem unele precizări despre rolul inteligenței artificiale în contextul disciplinelor cognitive.

Spectrul cercetărilor subsumate denumirii „Inteligență artificială” este foarte larg iar obiectivele proprii redeschid problemele celor mai importante domenii tradițional destinate modelării cognitive: percepția, raționamentul, limbajul și acțiunea. De altfel, ar fi multe de spus asupra acestei diviziuni care este inevitabilă, dar și foarte discutabilă, întrucât presupune posibilitatea unei separații efective și pare să propună o oarecare ordine, cronologică sau ierarhică, spre a le lua în considerație. Desigur că în fiecare din aceste *metadomenii*, problemele memoriei și ale învățării sunt fundamentale. Inspectând sumarele manualelor sau ale lucrărilor clasice despre inteligența artificială vom recunoaște un anumit număr de *teme relevante* pentru studiul cogniției: recunoașterea formelor și a cuvintelor, reprezentarea cunoștințelor, înțelegerea și producerea textelor, planificarea și diagnosticul etc.

Împrumuturile, ca și contribuțiile privind lingvistica sau neuropsihologia, care inspiră modelele conexioniste, nu sunt deloc rare. Iar în interiorul cercetărilor de inteligență artificială se întâlnesc efectiv lucrări autentic pluridisciplinare, unde informatica nu este decât un instrument privilegiat, servind modelării inspirate și controlate de discipline cu un *statut științific*, presupus a fi mai bine stabilit. Inteligența artificială oferă acestor *discipline ale cogniției* umane sau naturale ocazia de a-și corela obiectele specifice, de a-și extinde aria obișnuită de intervenție, și le mai oferă prilejul de a-și confrunta noțiunile descriptive cu concepțiile cele mai avansate de tratare a informației, care pot să beneficieze la rândul lor de aceste corelații și evoluții.

Mai curând decât o disciplină științifică omogenă, ceea ce ar presupune o anumită delimitare a obiectelor proprii, prin tradiție imposibilă, Inteligența Artificială poate fi considerată o *interdisciplină*, după cum spunea Herbert Simon încă de la începuturi. El dorea să afirme, astfel, că Inteligența Artificială nu se reduce numai la o serie de lucrări pluridisciplinare, că reprezintă mai mult decât o simplă arie problematică, unde cercetători în discipline separate se întâlnesc din când în când datorită unor proiecte tranzitorii. Cu toate că este, prin definiție, deschisă spre tot ceea ce înseamnă cugetare sau acțiune, Inteligența Artificială nu este, din acest motiv, mai puțin dotată cu un statut propriu, cu intenții și cu obiecte de predilecție. Atunci, în contextul științelor

cognitive, Inteligența Artificială deține rolul unei cutii de rezonanță epistemologică, fiind sensibilă la toate efectele de distorsiune și de simplificare ce se pot imagina. Ea profită, într-adevăr, de structurile descrise de alte discipline și caută, fără prea multă ezitare, prin experimentări și probe specifice, să le reconstruiască, să le regăsească, drept produse sau resurse ale anumitor procese calculatorii. Ea propune mereu reprezentări noi, proprii, care dobândesc adeseori valoare de normă în comparații metodice sau în transcrieri procedurale, nu numai în practică, ci și în teorie.

## 2. CONEXIONISMUL ÎN ȘTIINȚELE COGNIȚIEI

La începutul anilor '80 a avut loc un reviriment puternic al cercetărilor asupra rețelelor neuronale, considerate de atunci înainte ca fiind alcătuite din *neuroni formali*, spre a semnala îndeosebi faptul că nu mai era cazul de a menține perspectiva clasică, din neurobiologie. Revirimentul a fost ocazionat prin descoperirea noilor *tehnici de învățare* ce implicau diversele straturi de neuroni ascunși. În plus, trebuia menționată aplicarea pentru studiul acestor rețele a metodelor provenind din fizica sistemelor dezordonate, respectiv, tratarea în acest mod a comportamentului colectiv al atomilor dintr-un cristal, când este supus acțiunii unui câmp magnetic. Multitudinea proprietăților pe care aceste tehnici, împrumutate adeseori din termodinamică, le evidențiază în rețelele neuronale depășește tot ceea ce s-a realizat în anii precedenți. Iar aceste proprietăți au fost, curând, interpretate în termeni cognitivi și, astfel, au apărut formulări de genul: rețelele sunt capabile să învețe, să recunoască forme, să memoreze prin asociație etc. Reușita acestei tendințe în cercetare a avut ca efect aprecierea ei drept principala concurentă față de Inteligența Artificială ortodoxă, relaxată treptat, ca și când ar fi atins propria limită de competență.

Cea mai mare parte din controversele care mobilizează în prezent științele cognitive derivă din compararea celor două modele, succesiv promovate la rangul de paradigme. *Cognitivismul ortodox*, pentru care a gândi înseamnă a calcula ca și calculatorul, adică folosind simboluri ce au simultan realitate materială și valoare semantică (de reprezentare) este confruntat cu *paradigma rețelelor neuronale*, pentru care a gândi înseamnă a calcula așa cum procedează rețelele, funcționând masiv în paralel și unde comportamentele interesante apar numai la nivelul ansamblului, adică „emerg“ din sistemul de interacțiuni al unităților integrate.

De altfel, orice știință cognitivă, spre a-și justifica existența este obligată să propună o serie de modele, respectând imperativul coerenței sau consistenței logice, dar, totodată, să le valideze sau să le verifice prin confruntarea cu datele unor experiențe pertinente. În raport cu alte științe, diferența caracteristică disciplinelor cogniției presupune condiția ca datele experimentale să nu se refere direct la universul fizic, ci la modul în care indivizii umani și-l reprezintă. Într-adevăr, omul deliberază singur, alteleori discută cu alții, înainte de a hotărî asupra problemelor cu care se confruntă. Cu alte cuvinte, indivizii umani exprimă totul într-un limbaj, spre a comunica celorlalți uneori decizia și adesea chiar deliberarea respectivă. Există două stiluri de reprezentare semantică (modul realist și cel antirealist), cărora le sunt asociate teorii opuse cu privire la reprezentările mentale în Inteligența Artificială. „Disputa între realiști și antirealiști din *filosofia spiritului* interesează mai ales semantica sau forma logică a enunțurilor presupuse în raporturile atitudinale“ (D. H. Helman). Pentru un realist, interpretarea sau referința propoziției este o stare a universului și nicidecum o stare psihică (conștientă) a agentului reprezentând subiectul raportului de atitudine.

Se acceptă despre savanții cognitiști că apelează numai la criterii împrumutate din științele naturii (fizică, biologie etc.) și din cele tehnice sau ingineresti. Dar aprecierea performanțelor obținute de calculator datorită structurilor de programe „inteligente“, raportate la modalitățile de funcționare a spiritalului uman, deci nu numai la activitățile creierului, prezintă în acest domeniu o importanță hotărâtoare. Pornind de la o astfel de confruntare,

pot fi produse la un nivel adecvat de generalitate acele modele teoretice unitare, care să corespundă unor fenomene atât de eterogene. Una dintre problemele care a rămas deschisă privește contribuția informației figurative, adică a imaginilor mentale, la reprezentarea semantică rezultând din înțelegerea unui text. Unii autori au dezvoltat relativ la această problemă, o concepție asupra unui *model mental* (Johnson-Laird, A. Garnham). Însă ideea că acest model reprezintă o anumită stare de lucruri ar trebui să ne conducă, după alți autori, să discutăm mai curând despre un *model al situației* (Perrig și Kintsch). Toți acești autori au crezut util să deosebească în psihicul subiectului, alături de o „reprezentare propozițională” și o „reprezentare situațională”, dependentă de felul informației oferită agenților cognitivi. Pentru J.F.le Ny „reprezentările semantice tranzitorii pot fi considerate în mod rezonabil ca niște instanțieri ale reprezentărilor durabile, antrenând memoria pe termen lung” (1989, p. 221). Deoarece aceste instanțieri sau concretizări sunt realizate la om printr-un proces de activare, se poate spune că a *instanția* înseamnă a deveni activ. În acest sens, autorul citat apreciază că, raportat la om, sistemul de calcul implementat într-un ordinator „nu activează altfel, ci doar cu mijloace fizice foarte diferite”.

Conexionismul recent își pune, în privința rețelelor considerate, problema reprezentării simbolice a „universului exterior”. Dar, întrucât astăzi cele mai multe ordinatoare sunt încă secvențiale, motiv pentru care sunt denumite „mașini Von Neumann”, chiar rețelele neuronale din lucrările conexioniștilor, aceste ansambluri multiplu paralele, sunt simulate pe astfel de ordinatoare secvențiale! Problema complexității este aproape eludată când se reduce o funcție la o structură. Problema cunoașterii capacității (puterii) unei structuri complexe se apreciază astfel: fiind dată o clasă de structuri definite numai parțial și în mod statistic, se încearcă explorarea posibilităților generale sau generice. Modelele în rețea au fost inițial admise ca macro-modele; nodurile, ca și arcele componente, erau considerate purtătoare ale unor *unități cognitive* de nivel înalt, adică simbolice sau conceptuale. Dar modelele neoconexioniste se bazează pe rețele alcătuite din microunități; nodurile și arcele componente nu mai sunt denumite individual și nu mai sunt considerate ca deținătoare sau transmițătoare ale unor pachete organizate de



informații. Între cele două tipuri de abordare (conexionism și neoconexionism) trebuie văzută diferența care provine din nivelurile deosebite de cercetare a fenomenelor mentale. Modelele numite *simbolice*, spre deosebire de cele alcătuite din rețele *subsimbolice*, sunt tocmai acelea care se raportează la extensiuni sau intensiuni ale reprezentărilor conceptuale, determinate de universul exterior.

Putem spera îndeosebi că prin connexionism vom avea explicată capacitatea unei rețele de a produce logica, această capacitate fiind o *proprietate emergentă* a rețelei luată în ansamblu. Atunci, izvorul capacității noastre de a gândi logic nu poate fi localizat la un nivel inferior și nici considerat ca dependent numai de elemente izolate. Iar faptul că logica a fost utilizată spre a modela rețeaua neuronală nu presupune că rețeaua ca atare ar putea fi modelul gândirii logice. Mai rămâne de specificat maniera în care semnalul emis de un anumit neuron poate reprezenta o propoziție asupra lumii, modalitatea de codificare sau de echivalare a unui conținut de gândire cu calculul operat în rețea și, respectiv, modul în care acest cod poate dobândi o semnificație sub formă de reprezentare. Dacă o stare mentală reprezintă ceva numai în măsura în care prezintă un conținut, acest conținut trebuie să se refere la universul fizic. Capacitatea reprezentativă a stărilor mentale se numește *intenționalitate* și constă în faptul că se raportează la anumite obiecte eterogene.

În orice caz, nu se mai revine la ideea perimată a psihologiei asociaționiste, conform căreia structura mesajului receptat determină total structura mesajului de ieșire. Fără ca sistemul nervos să fie considerat izolat față de universul exterior, semnalul de la intrare nu mai este privit decât ca un declanșator, care conduce la o alegere între diversele moduri posibile de funcționare autonomă a sistemului nervos și care poate impune schimbarea modului de funcționare anterioară. În afara contactului cu fizica sistemelor dezordonate nu se putea concepe varianta neoconexionismului. O rețea aleatorie poate învăța și în particular poate efectua inferențe inductive, dar este nevoie de ceva în plus spre a deveni capabilă să exprime regulile pe care le urmează în activitatea proprie. O regulă nu se poate modifica decât prin intervenția unei metareguli. Dar neoconexionismul se raportează la rețea ca la un obiect complex, în sensul lui Von

Neumann: cvasinatural. Astfel, i se pot studia unui asemenea obiect în mod *experimental* proprietățile, fiind manipulat, adică variindu-se parametrii săi. Însă nu se poate accepta punctul de vedere al totalei cunoașteri de către ordinator a proprietăților sale.

S-a observat că orice sistem compus dintr-un foarte mare număr de elemente ce interacționează dezvoltă inevitabil și de la sine proprietăți de stabilitate. De exemplu, automatele booleene conectate în mod aleator generează un comportament colectiv remarcabil de tip ciclic și cu o perioadă redusă. S-a mers prea departe cu autonomizarea modelului logic-matematic, cu toate că se pretinde o menținere a legăturii cu realitatea fizică, biologică și chiar psihologică. Reconstrucția conceptuală a subiectului a evidențiat numeroasele riscuri de eroare pe care le comportă astfel de încercări. Noua „știință a populațiilor“ referitoare la sisteme automate sau moleculare abandonează una din categoriile fundamentale pe care s-a clădit dinamica, inclusiv termodinamica clasică: funcția de stare, deoarece fenomenele colective pe care le studiază nu sunt integrabile. Încercând să depășească atât cauzalitatea lineară cât și cauzalitatea circulară, această *nouă știință* s-a angajat într-o „cauzalitate de cuplări“ (I. Stengers), unde nu se specifică natura interacțiunilor. Dar dacă vrem să regândim intenționalitatea ca raport între o stare mintală și obiectul său (conținut), atunci trebuie să ne rupem de fizică și chiar de noua știință ce racordează conexiionismul la fizica sistemelor dezordonate și la termodinamica proceselor ireversibile.

### **3. EXPLICAREA MENTALULUI SAU EXPERIENȚA ÎN COGNITIVISM**

Mai degrabă decât „gândirea înseamnă calcul“, formula care ar trebui să rezume directiva științelor cognitive ar putea fi: „cunoașterea presupune simularea“. Deși utilizat în mod asemănător cu „modelarea“, termenul „simulare“ prezintă semnificații diferite, eterogene, după cum apare încadrat într-un

context științific sau este considerat în accepția curentă. În primul caz, *simularea* reprezintă o formă particulară de modelare ce constă în reproducerea pe calculator a parametrilor funcționării unui sistem oarecare; deci, suntem în domeniul activităților cognitive. În al doilea caz, simularea presupune o *disimulare*; adică suntem în domeniul simulacrului, al imitației sau al prefacerii (acțiunii). V. Descobes (1979) supralicitează, presupunând că „orice simulare admite a simula simularea“. Așadar, unii teoreticieni ai Inteligenței Artificiale acceptă o dublă prefăcătorie: discutăm ca și când calculatorul se prefacă că știe să calculeze. În acest sens, *cotitura mentalistă* apărută prin „știința spiritului“ a propus o deconstrucție a metafizicii subiectivității care conduce mult mai departe decât demersul filosofic omonim.

A reprezenta, a imita, a repeta sau a reproduce: iată componente permanent solicitate în orice *experiență științifică*. Dar există ceva în activitatea științifică aspirând spre universalul ce scapă experimentării, deși confirmă perfect analizele omului de știință: construcția de modele. Știința este activitatea umană în contextul căreia termenul „model“ are un sens diametral opus celui folosit în limbajul cotidian, unde *model* înseamnă doar ceva de imitat sau care merită să fie imitat. Paradoxal, însuși modelul științific a fost la origine o imitație, având cu realitatea același raport ca o miniatură (modelul redus) cu obiectul pe care îl reproducem spre a-l manipula mai ușor. Ulterior, modelul va fi înțeles ca o formă abstractă, care se încorporează sau se realizează în anumite fenomene luate în discuție. Domenii foarte diferite ale realității fenomenale pot fi reprezentate folosind modele indentece.

Deoarece astfel se stabilește între domeniile respective o relație de echivalență se spune adeseori că modelul este clasa echivalenței corespunzătoare. Întrucât și modelul în sens științific reprezintă o creație umană, există o anumită supra-punere între cel ce imită și ceea ce se imită. Cu alte cuvinte, prin modelarea științifică obținem o imitare a naturii pe care savantul o consideră mai degrabă un model în sensul obișnuit. Conceperea și înfăptuirea modelelor conferă omului o dominație pe care el nu o poate dobândi prin cunoașterea directă, intuitivă sau empirică. Această dominație presupune nu numai explicația, dar și predicția sau anticiparea, îndeosebi datorită puternicului instrument matematic utilizat.

După J. Fodor (1975), *teoria reprezentățională* a mentalului se justifică printr-o lungă serie de lucrări din psihologia cognitivă și întemeiază orice explicație corectă a cogniției. Conform tezei funcționaliste a lui Fodor, tipurile de stări mentale, cum ar fi credințele și dorințele, reprezintă *stări funcționale* ale organismului uman care pot fi definite prin relațiile lor funcționale și cauzale cu stimuli, cu manifestări comportamentale sau cu alte stări mentale. Relația cauzală dintre conținuturile semantice ale reprezentărilor și manifestările comportamentale poate fi mai ușor înțeleasă dacă recurgem la *metafora ordinatului*, deoarece ordinatoarele au tocmai însușirea de a lega proprietățile cauzale ale simbolurilor cu proprietățile lor semantice prin intermediul structurii lor sintactice. Dacă spiritul este un fel de ordinator, putem spune, după Fodor, că avem asigurată o soluție la *problema cauzalității*: inscripția concretă a simbolurilor, individualizate prin forma lor, leagă proprietățile semantice ale reprezentărilor, adică faptul că ele se referă la ceva anume, cu *stările fizice* ale mașinii electronice (J. Fodor, 1987.). O *explicație cognitivă* este chiar o explicație cauzală, întrucât postulează secvențe cauzale de reprezentări simbolice incluse în *relații* computaționale. Z. Pylyshyn (1984) recunoaște că un comportament este cognitiv „penetrabil” dacă se conformează unei explicații de acest tip. Totodată, aceste reprezentări simbolice și relații computaționale sunt interne, iar o psihologie științifică trebuie să se supună principiului numit de Fodor „solipsism metodologic”. Acest principiu respinge orice individualizare a conținuturilor mentale în termeni care s-ar referi la condiții externe psihologiei unui subiect și în special care nu ar fi determinate prin alcătuirea sa și prin legile sale de organizare materială. Mai precis, *principiul solipsismului metodologic* ne impune să recurgem doar la o individualizare sintactică a conținuturilor mentale considerate simboluri concrete și nu la o individualizare semantică, adică folosind referiri la entitățile lumii exterioare.

Sistemele informatice destinate unei automatizări parțiale a raționamentelor sunt cunoscute sub numele de *sisteme expert*. Ele au devenit curând „revelatoare pentru un anumit număr de probleme care privesc nu numai Inteligența Artificială, dar și ansamblul tehnostiințelor cogniției” (Y. -M. Visetti, 1993). Referința la sistemul expert, prin comparație cu intelectul uman,

are ca efect să evidențieze o dată în plus diferența care există între conceptualizarea fenomenelor naturale și cea a fenomenelor raționale. De exemplu, diferența care se stabilește între: 1) *relieful momentan* pe care, la un subiect uman, îl poate lua un aspect, de la un moment anumit, în raport cu reprezentarea totală a unei situații date (atenția sau importanța pe care intelectul acestui subiect o acordă în mod spontan unui aspect particular al reprezentării de ansamblu) și 2) *pertinența*, într-un anumit sens obiectivă, a unei proprietăți dintr-o situație reală. Sunt două probleme bine determinate și distincte să indentificăm ceea ce este pertinent într-o situație și ceea ce este *tratat drept* pertinent, adică scos în relief, la un moment dat, de către un individ aflat, efectiv sau numai fictiv, în această situație considerată. Fără îndoială că există o relație generală de corespondență între pertinența obiectivă și relieful momentan. În versiunea sa optimistă, această relație de corespondență poate fi formulată în modul următor: în general, subiectul acordă atenție și relief proprietăților obiectiv pertinente ale situațiilor. Dimpotrivă, în versiunea pesimistă, se evidențiază mai degrabă discordanțele relației.

Aceeași relație poate fi considerată însă ca fiind o relație de dependență mai curând decât de corespondență. Atunci, ea este întemeiată pe învățare sau exercițiu. Formularea adecvată va fi, în acest caz, următoarea: în general, subiectul învață progresiv să acorde mai mult relief și atenție proprietăților pertinente ale situațiilor. Pe o astfel de formulare se poate baza și o acțiune didactică eficace. Putem adăuga că învățarea de acest gen nu rămâne exclusiv individuală. Sarcina istorică a științei este tocmai de a identifica proprietățile lucrurilor care dețin în mod obiectiv cea mai mare pertinență și importanță în Univers. Iar sarcina socială și individuală a învățământului este de a urmări ca aspectele/atributele, care vor avea mai mult relief în reprezentările individuale, să fie tocmai cele care corespund în mod efectiv proprietăților obiectiv pertinente sau importante.

În concluzie, metoda experimentului mental nu mai conține prejudecata empiristă sau verifiționistă, atât de răspândită. Într-adevăr, metoda experiențelor mentale din filosofia analitică s-a dovedit fecundă în măsura în care a fost însoțită de reflecția critică. Metoda *cazurilor imaginare* nu poate fi contestată, chiar dacă „variațiile imaginare“ se raportează la trăsături fizice sau fiziologice pentru a determina criteriile de identitate a subiec-

ților. De astfel, conceptele noastre nu pot fi clarificate decât prin supunerea lor la astfel de variații și, în final, prin trecerea la limita lor de aplicabilitate.

„Studiile“

## NOTE

<sup>1</sup> J. M. Besnier, „L'intelligence artificielle et les paradoxes de la modernité“, p. 52 în *Les nouveaux outils du savoir*, culegere îngrijită de P. Chambart și P. Lévy, Ed. Descartes, Paris, 1993.

<sup>2</sup> G. Boss, *Les Machines à penser: l'Homme et l'ordinateur*, p. 152, Ed. du Grand Midi, Zürich, 1987;

<sup>3</sup> D. C. Dennett, „Ou suis-je?“ în *Vues de l'esprit*, în colab. cu D. Hoffstadter, InterEditions, 1987;

<sup>4</sup> A. Garnham, „Mental models as representations of text“, p. 560 în *Memory and cognition*, ON 9/1990;

<sup>5</sup> D. H. Helman, „Realism and Antirealism“ în A. I. în *Brit. J. Phil. Sci.* 24/1990;

<sup>6</sup> P. N. Johnson-Laird, *Mental models*, 1983, Cambridge Univ. Press.

<sup>7</sup> J. von Neumann și N. Wiener, *From Mathematics to the Technologies of Life and Death*, Mit Press, Cambridge, 1980;

<sup>8</sup> J. F. Le Ny, *Science cognitive et compréhension du langage*, P. U. F. Paris 1989;

<sup>9</sup> D. Parisi, *Intervista sulle reti neurali; cervello e macchina intelligenti*, Bologna, Il Mulino, 1989;

<sup>10</sup> W. Perrig și W. Kintsch, „Propositional and situational representations of text“, în *Journal of Memory and Language* 24/1985, p. 333;

<sup>11</sup> H. Simon, *Models of my Life*, Basic Books, 1991;

<sup>12</sup> I. Stengers, „Les généalogies de l'auto-organisation“, în *Cahiers du creation*, 8/nov. 1985;

<sup>13</sup> T. Trabasso și L. L. Sperry, „Causal relatedness and importance of story events“, în *Journal of Memory and Language*, 24/1985;

<sup>14</sup> V. Descomber, *Le Même et l'Autre. Quarantecinq ans de philosophie française*, 1979, Minuit, Paris;

<sup>15</sup> P. Engel, *La norme du vrai, philosophie de la logique*, 1989, Gallimard, Paris;

<sup>16</sup> J. Fodor, *The Language of Thought*, MIT Press, Cambridge Mass. 1975;

<sup>17</sup> J. Fodor, *Psychosemantics, The Problem of Meaning in the Philosophy of Mind*, MIT Press, 1987;

<sup>18</sup> S. Kosslyn, *Image and Mind*, 1980, Harvard U. P.;

<sup>19</sup> P. Johnson-Laird, *Mental Models*, 1983, Cambridge U. P.;

<sup>20</sup> Pylyshyn, *Computation and Cognition*, 1984, MIT Press, Cambridge;

<sup>21</sup> R. Shepard & L. Cooper, *Mental Images and their Transformations*, MIT Press, 1982, Cambridge Mass.; Y. -M. Visetti.;

<sup>22</sup> „Intelligence artificielle et systèmes experts: quelques remarques épistémologiques“ în vol. *Les nouveaux outils du savoir*, 1993, Editions Descartes. Paris 1993.

<sup>23</sup> F. Yaters, *L'art de la mémoire*, 1980, Gallimard, Paris.

# BIBLIOGRAFIE SELECTIVĂ

- Alexander, R. D: *The Biology of Moral Systems*, (Aldine de Gruyter, New York, 1987).
- Anscombe, G. E. M., *Intentions*, 2nd ed (Oxford: Blackwell, 1963).
- Anscombe, G. E. M., *An Introduction to Wittgenstein's Tractatus*, (Londra: Hutchinson, 1967).
- Anscombe, G. E. M., and Geach, P., *Three Philosophers: Aristotel, Aquinas and Frege* (Oxford; Blackwell, 1961).
- Audi, Robert, *Belief, Justification, and Knowledge: An Introduction to Epistemology* (Belmont, CA: Wadsworth Publishing Company, 1988).
- Austin, J. L., *Sense and Sensibilia*, reconstituire de G. J. Warnock (Oxford: University Press, 1962).
- Ayer, A. J., *Philosophy in the Twentieth Century* (Londra: Weidenfeld & Nicolson, 1982; Unwin Paperbacks, 1984).
- Ayer, A. J., *The Problem of Knowledge* (Harmondsworth: Penguin, 1956).
- Ayer, A. J. and Winch, P., *British Empirical Philosophers: Locke, Berkeley, Hume, Reid, and Mill* (Londra: Routledge, 1952).
- Beakley, B., și Ludlow P., (ed), *The Philosophy of Mind (Classical Problems/Contemporary Issues)*, (MIT Press, Cambridge, SUA, 1992).
- Blackburn, S., *Oxford Dictionary of Philosophy*, Oxford Univ. Press, Oxford, 1995.
- Blakemore, C., *Mechanisms of the Mind*, B. B. C. Reith Lectures 1976 (Cambridge: University Press, 1977).
- Block, Ned „Troubles with Functionalism“ în W. Savage (ed.), *Perception and Cognition: Issues in the of Psychology*, Minnesota Studies in the Philosophy of Science, Vol. IX (Minneapolis, MN: University of Minnesota Press, 1978)
- Block, Ned „Introduction: „What is Functionalism?““ în N. Block (ed.), *Readings in Philosophy of Psychology*, Vol. 1 (Cambridge, MA: Harvard University Press, 1980).
- Boden, M. A., *Minds and Mechanisms* (Brighton: Harvester, 1981).
- Boden, M. A., *Artificial Intelligence and Natural Man*, (Londra, MIT Press, 1987).
- Boden, M. A., *Artificial Intelligence in Psychology. Interdisciplinary Essays*, (MIT Press, Londra 1989).
- Bogdan, Radu I. (ed) *Belief. Form, Content and Function*. (Clarendon Press, Oxford, 1986).
- Braden, St. E. *First Person Plural, (Multiple Personality and the Philosophy of the Mind)* (Routledge, Londra, 1991).
- Brand, Myles, *Intending and Acting* (Cambridge, MA: The MIT Press/Bradford 1984).

- Brandt Johannes și Wolfgang L. Gombocz (eds), *The Mind of Donald Davidson*, (Editions Rodopi B.V. Amsterdam, Atlanta, GA 1989)
- Brentano, Franz, *Descriptive Psychology*, trad. și reeditare: Benito Mullerr (Routledge, Londra, 1995).
- Brentano, Franz, *Psychology from an Empirical Standpoint*, trad. și reeditare: A. Rancurello, D.B. Terrell și L.L. Mcalister, (Routledge, Londra, 1995).
- Brown, James Robert, *The Laboratory of the Mind (Thought Experiments in the Natural Sciences)* (Routledge, Londra, 1991).
- Brzezinski, Jerzy, ed. (I) *Probability in Theory-Building (Experimental and Non-Experimental Approaches in Scientific Research in Psychology)* (Rodopi, Amsterdam, Atlanta, GA, 1994).
- Carruthers, Peter, *Introducing Persons (Theories and Arguments in the Philosophy of Mind)*, (Routledge, Londra, 1986).
- Chomsky, N., *Language and Mind* (New York: Harcourt Brace Jovanovich, 1972).
- Chomsky, N., *Reflections on Language*, (Londra: Collins/Fontana, 1976).
- Churchland, P.M., *Matter and Consciousness. A Contemporary Introduction to the Philosophy of Mind* (Cambridge Mass: MIT Press, 1984).
- Churchland, P. M., *Scientific Realism and the Plasticity of Mind*, (Cambridge University Press, Cambridge, 1979).
- Churchland, P.M., „Eliminative Materialism and Propositional Attitudes“, în *Journal of Philosophy*, vol. LXXVIII, 1981.
- Churchland, P. M., *A Neurocomputational Perspective (The Nature of Mind and the Structure of Science)* (MIT Press, Cambridge, SUA, 1989).
- Churchland, P.M., *The Engine of Reason, The Seat of the Soul (A Philosophical Journey into the Brain)* (MIT Press, Cambridge, SUA, 1995).
- Churchland, P. S., *Neurophilosophy towards a Unified Theory of Mind-Brain*, (MIT Press, Cambridge, 1986).
- Clark, Austen, *Psychological Model and Neural Mechanisms (An Examination of Reductionism in Psychology)*, (Clarendon Press, 1989).
- Clark, Andy, *Microcognition, Philosophy, Cognitive Science and Parallel Distributed Processing* (MIT. Press, Cambridge, SUA, 1989).
- Cohen, P. R., J. Morgan și M. E. Pollock, *Intentions in Communication*, (MIT. Press, Londra, 1990).
- Cohen, Simon Baron, *Mindblindness (An Essay on Autism and Theory of Mind)*, MIT Press, Cambridge, SUA, 1995).
- Collins, Harry M., *Artificial Experts. Social Knowledge and Intelligent Machines*, (MIT Press Londra, 1990).
- Coudill, Maureen and Charles Butler, *Naturally Intelligent Systems*, (MIT Press, Londra 1990).
- Crane, Tim, *The Mechanical Mind (A Philosophical Introduction to Minds, Machines and Mental Representation)*, (Penguins Book, 1995).
- Crimmins, Mark, *Talk abot Beliefs* (MIT Press Cambridge, SUA, 1992).
- Cummins, Robert C., *The Nature of Psychological Explanation* (Cambridge, MA: The MIT Press/Bradford) (1983).
- Cummings, Robert, *Meaning and Mental Representation* (MIT Press, Cambridge, SUA, 1989).



- Cummings, R. și J. Pollock, *Philosophical AI Essays at the Interface* (MIT Press, Cambridge, SUA, 1991).
- Davidson, D., *Mental Events* în Foster, L., și Swanson, J. W. (eds.) *Experience and Theory*, (University of Massachusetts Press, Amherest, 1970).
- Davidson, D. *Inquiries into truth and interpretation*, (Oxford Univ. Press, Oxford, 1984).
- Dawkins, R., *The Blind Watchmaker* (Harmondsworth: Penguin, 1988).
- Dawkins, R., *The Selfish Gene* (Oxford: University Press, 1986).
- Dennett, D.C., *Brainstorms. Philosophical Essays on Mind and Psychology* (Brighton: Harvester, 1978).
- Dennett, D.C., „Three Kinds of Intentional Psychology“, în Healey, R. (ed.) *Reduction, Time and Reality* (Cambridge University Press, Cambridge, 1981).
- Dennett, D. C., *The Intentional Stance* (MIT Press, Cambridge, SUA, 1987).
- Dennett D. C., *Content and Consciousness*, (Routledge, Londra, 1989).
- Diamond, Cora, *The Realistic Spirit, Wittgenstein Philosophy and Mind*, (MIT Press Cambridge, SUA, 1991).
- Dretske, Fred, *Explaining Behavior (Reason in a World of Cause)* (MIT Press, Cambridge, 1988).
- Dummett M. *Elements of Intuitionism*, (Clarendon Press, Oxford, 1988).
- Eccles, John C., *The Human Psyche*, (Routledge, Londra, 1992).
- Eccles, John C., *Evolution of the Brain: Creation of the Self*, (Routledge, Londra, 1991).
- Eckart, Barbara von, *What is Cognitive Science*, (MIT Press, Cambridge, SUA, 1992).
- Finke, Ronald A., *Principles of Mental Imagery*, (MIT. Press, Londra, 1989).
- Fiumara, Gemma Corradi, *The Metaphoric Process (Connections Between Language and Life)* (Routledge, Londra, 1979).
- Flanagan, Owen, *Consciousness Reconsidered* (MIT. Press, Cambridge, SUA, 1992).
- Fodor, Jerry A. *The Language of Thought* (Cambridge, MA: Harvard University Press), 1975.
- Fodor, Jerry A., „Methodological Solipsism Considered as a Research Strategy in Cognitive Psychology“, *The Behavioral and Brain Sciences*, 3, 1980.
- Fodor, Jerry A., „The Present Status of the Innateness Controversy“ în J. A. Fodor, *Representations: Philosophical Essays on the Foundations of Cognitive Science* (Cambridge, MA: The MIT Press/Bradford), 1981.
- Fodor, Jerry A., *The Modularity of Mind: An Essay on Faculty Psychology* (Cambridge, MA: Thea MIT Press/Bradford, 1983).
- Fodor, Jerry A., *Psychosemantics: The Problem of Meaning in the Philosophy of Mind* (Cambridge, MA: The MIT Press/Bradford), 1987.
- Fodor, Jerry A. Z. W. Pylyshyn, „Connectionism and Cognitive Architecture: A Critical Analysis“, *Cognition*, 28, 1988.
- Fodor, Jerry A., *A Theory of Content and Other Essays*, (MIT Press, Cambridge, SUA, 1990).
- Fodor, Jerry A. *The Elm and the Expert (Mentalese and its Semantics)*, (MIT Press Cambridge, SUA, 1994).

- Foster, John, *The Immaterial Self (A Defence of the Cartesian Dualist Conception of the Mind)* (Routledge, Londra, 1991).
- Foucault, M., *The Archaeology of Knowledge*, (New York, Harper and Row, 1975).
- Fraassen, B. Van, *The Scientific Image*, (Cambridge University Press, Cambridge, 1980).
- Gärdenfors, Peter, *Knowledge in Flux. Modelling the Dynamics of Epistemic States*, (Mit Press Londra, 1990).
- Gardner, S. *Irrationality and the Philosophy of Psychoanalysis*, (Cambridge Univ. Press, Cambridge, 1993).
- Garfield, Jay L., *Belief in Psychology. A Study in the Ontology of Mind*. (Mit. Press Londra, 1988).
- Garnham, Alan, *Artificial Intelligence. An Introduction*, Routledge, Londra 1988.
- Gavroglu, K., John Stachel, M. W. Wartofsky (ed), *Science, Mind and Art*, In onoarea lui R.S. Cohen, seria Boston Studies in the Philosophy of Science 165, (Kluwer Acad. Publ., Dordrecht, 1988).
- Glover, J. (ed.), *The Philosophy of Mind* (Oxford: University Press, 1976; Oxford Readings series).
- Glymour, Clark, *Thinking Things Through. An Introduction to Philosophical Issues and Achievements* (Mit Press, Cambridge-SUA, 1992)
- Goldman, Alvin I. (ed), *Reading in Philosophy and Cognitive Science* (Mit Press, Cambridge, SUA, 1993).
- Gregory, R., L., (ed), *The Oxford Companion to the Mind*, Oxford Univ. Press, Oxford, 1987.
- Grene, M., *Descartes* (Brighton: Harvester, 1985; Philosophers in Context series).
- Hacker, P.M.S., *Insight and Illusion: Wittgenstein on Philosophy and the Metaphysics of Experience* (Oxford; University Press, Oxford, 1972).
- Hacker, P.M.S. *Appearance and Reality* (Blackwell, Oxford, 1987).
- Hacking, I., *Representing and Intervening*, (Cambridge University Press, Cambridge, 1983).
- Hamlyn, D.W., *The Theory of Knowledge* (Londra: Macmillan, 1971).
- Hampshire, S., *Thought and Action*, (Londra: Chatto and Windus, 1982).
- Hare, R.M., *Freedom and Reason* (Oxford: University Press, 1965).
- Harré, R., *The Philosophies of Science: An Introductory Survey*, 2nd ed (Oxford: University Press, 1985).
- Harré, R., „Wittgenstein and Artificial Intelligence“ în *Philosophical Psychology*, vol. 1, nr. 1/1988.
- Hick, J., *Faith and Knowledge* (Ithaca, New York: Cornell Univ. Press, 1967).
- Honderich, Ted, *How Free are You? The Determinist Problem*, (Oxford Univ. Oxford, Press, 1993).
- Hookway, C. (ed.), *Minds, Machines and Evolution: Philosophical Studies* (Cambridge: University Press, 1984).
- Hyslop Alee, *Other Minds*, (Kluwer Academic Publ., Dordrecht, 1995).
- Jackendoff, Ray, *Semantics and Cognition*, (Mit Press, Londra, 1989).
- Jackendoff, Roy, *Consciousness and the Computational Mind* (Mit Press, Londra, 1991).

- Jackendoff, Ray, *Languages of the Mind*, (Mit. Press, Cambridge, SUA, 1992).
- Johnson-Laird, P.N., *Computers and the Mind* (Cambridge: Harvard University Press; si Londra: Collins/Fontana, 1988).
- Keil, Frank C., *Concepts, Kinds, and Cognitive Development*, (Mit Press, Londra, 1989).
- Kenny, A., *Wittgenstein* (Harmondsworth: Penguin, 1975).
- Kitcher, Patricia, *Freud's Dream, A Complete Interdisciplinary Science of Mind*, (Mit Press, Cambridge, SUA, 1992).
- Kornblith, Hilary (ed), *Naturalizing Epistemology* (ed. II-a) (Mit. Press, Cambridge, SUA 1993)..
- Lacan, Jaques, *Ecrits. A Selection*, trad. A. Sheridan, (Routledge, Londra, 1980).
- Langer, S.K., *Philosophy in a New Key*, 3rd ed (Cambridge, Mass: Harvard University Press, 1977).
- Layton, Michall, *Symmetry, Causality, Mind*, (Mit Press, Cambridge SUA, 1992).
- Lepore, H., McLoughlin (eds), *Actions and Events*, (Oxford, Basil Blackwell, 1985).
- Levin, M. E., *Metaphysics and Mind-Body Problem*, (Clarendon Press, Oxford, 1988).
- Lewis, H.D., *Teach Yourself: Philosophy of Religion* (Londra, English Universities Press, 1965).
- Lorenz, K., *Behind the Mirror*, trad. R. Taylor (Londra, Methuen, 1977).
- Lumsden, C., Wilson, E., *Genes, Mind and Culture: the Coevolutionary Process*, (Cambridge Mass, Cambridge, 1981).
- Lumsden, C., Wilson, E., *Prometheus Fire: Reflection of the Origin of Mind*, (Cambridge Mass, Cambridge, 1983).
- Lycan, William G., *Consciousness*, (Mitt Press, Cambridge SUA 1987).
- Lyons, W., *Gilbert Ryle: An Introduction to this Philosophy* (Brighton: Harvester, 1980).
- Mac Cormae, E. B., *A Cognitive Theory of Metaphor*, (Mit Press, Cambridge, SUA, 1989).
- MacDonald, Cynthia, *Mind-Body Identity Theories*, (Routledge, Londra, 1992).
- MacIntyre, A.C., *The Unconscious: a Conceptual Analysis* (Londra, Routledge, 1958).
- Manktelow, K. și D.E.Over, *Rationality, Psychological and Philosophical Perspectives*, (Routledge, Londra, 1993).
- McCulloch, Warren S., *Embodiment of Mind*, (Mit Press, Londra, 1988).
- McGinn, C., *The Character of Mind* (Oxford: University Press, 1982).
- Mellor, D. H., *Nothing like Experience*, P.A. S 93, 1992.
- Merleau-Ponty, M., *The Phenomenology of Perception*, trad. C. Smith (Londra: Routledge, 1962).
- Merleau-Ponty, M. *The Structure of Behaviour*, trans. J. Wild (Londra: Methuen, 1955).
- Millikan, Ruth G., *Language, Thought and Other Biological Categories* (New Foundation for Realism), Cambridge, MA: MIT. Press, 1984.
- Moser, Michael C., *The Perception of Multiple Objects (A Connectionist Approach)* (Mit Press. Londra. 1991).
- Mundle, C.W.K., *Perception: Facts and Theories* (Oxford: University Press, 1971).

- Osherson, Daniel N., *An Invitation to Cognitive Science*. (Mit Press, Londra, 1990).
- Osherson, Daniel N., St. M. Kosslyn și John Hollerbach (eds), *Visual Cognition and Action*, (Mit Press, Londra, 1990).
- Osherson, Daniel N. și E.E. Smith (eds), *Thinking*, (Mit Press, Londra, 1990).
- Ornstein, Robert E. *The Psychology of Consciousness* (San Francisco, CA: W. H. Freeman and Company, 1972).
- Ornstein, Robert E., *The Nature of Human Consciousness: A Book of Readings* (San Francisco, C. A.: W. H. Freeman and Company, 1973).
- Papineau, D., *Philosophical Naturalism* (Oxford: Basil Blackwell, 1993).
- Papineau, D., „Why Supervenience?“ *Analysis*, 50, 1990.
- Papineau, D., „The Reason Why“, *Analysis*, 51, 1991.
- Peacocke, C., *Holistic Explanation*, (Oxford: Clarendon Press, 1979).
- Peacocke, Christopher, *A Study of Concepts*, (Mit Press, Cambridge, SUA, 1992).
- Parfit, D., *Reasons and Persons* (Oxford: University Press, 1984).
- Penrose, Roger, *The Emperor's New Mind: Concerning Computers, Minds and the Laws of Physics*, (Penguin Books, Oxford University Press, 1989).
- Perner, Josef, *Understanding the Representational Mind*, (Mit Press, Londra, 1991).
- Pettit, Rh., „A Definition of Physicalism“, in *Analysis*, vol 53, 1993.
- Pollock, John, *How to Build a Person. A Prolegomenon*. (Mit Press, Londra 1989).
- Polanyi, M., *Personal Knowledge*, Univ Press, Chicago, 1958.
- Popper, K.R. și J.C. Eccles, *The Self and Its Brain. An Argument for Interactionism*, (Springer Int. Ed, Berlin, New York, 1977, Routledge, Londra, 1984).
- Popper, K.R., *Conjectures and Refutations: The Growth of Scientific Knowledge*, (Londra: Routledge, 1972).
- Popper, K.R., *Objective Knowledge. An Evolutionary Approach*, (Oxford: University Press, 1979).
- Posner, Michael J., *The Foundations of Cognitive Science*, The MIT Press, Londra, 1989.
- Putman, H., *Mind, Language and Reality: Philosophical Papers*, vol. 2 (Cambridge: University Press, 1975).
- Putman, H., *Meaning and the Moral Sciences* (Londra: Routledge and Kegan Paul, 1978).
- Putman, H., *Reason, Truth and History* (New York, NY: Cambridge University Press, 1981).
- Pylshyn, Zenon W., *Computation and Cognition, Toward a Foundation for Cognitive Science* (Mit. Press Londra 1986).
- Quine, W.V.O., *Word and Object* (Cambridge, Mass: MIT Press, 1960).
- Quine, W. *Ontological Relativity and Other Essays*, (Columbia University Press, New York et Londra, 1969).
- Raffman, Diane, *Languages, Music and Mind*. (Mit Press, Cambridge SUA, 1993).
- Recher, N., *Forbidden Knowledge and Other Essays on the Philosophy of Cognition* (Dordrecht, Kluwer Academic Publishers, 1985).
- Rorty, R., *Philosophy and the Mirror of Nature*, (Princeton, Princeton Univ. Press, 1979).

- Rorty, R. *Consequences of Pragmatism*, Univ of Minnesota Press, Minneapolis, 1982.
- Rorty, R., „Mind-Body Identity, Privacy and Categories“, în *Revue of Metaphysics* 19/1965.
- Rorty, R., Objectivity, Relativism and Truth, în *Philosophical Papers*, vol I, Cambridge Univ. Press, 1991.
- Russell B., *Human Knowledge: Its Scope and Limits* (Londra, Allen and Unwin, 1948).
- Ryle, G., *The Concept of Mind* (1949) (Harmondsworth: Penguin, 1963).
- Ryle, G., (ed.) *The Revolution in Philosophy* (Londra: Macmillan, 1956).
- Schrag, Calvin P., *Experience and Being*, Northwestern Univ. Press, Evanston, 1969.
- Schiffer, St., „Compositional Supervenience Theories and Compositional Meaning Theories“, în *Analysis* vol. 53, 1993.
- Scruton, R., *A Short History of Modern Philosophy from Descartes to Wittgenstein* (Londra: Routledge, 1981; Ark Paperbacks, 1984).
- Searle, J.R., *Intentionality: An Essay in the Philosophy of Mind*, (Cambridge: University Press, 1983).
- Searle, J.R., *Minds, Brains and Science* (Reith Lectures, 1984) (Harmondsworth: Penguin, 1989).
- Searle, J.R., *The Philosophy of Language* (Oxford, University Press, 1971).
- Searle, J.R., *Speech Acts* (Cambridge: University Press, 1969).
- Searle J. R. „Analytic Philosophy and Mental Phenomena“ în *Historical Foundation of Cognitive Science*, (Kluwer Acad. Publ., Boston 1990).
- Searle J. R., *The Rediscovery of the Mind*, (Mit Press, Cambridge, SUA, 1992).
- Sellars, W., „Empiricism and the Philosophy of Mind“, în *Science. Perception and Reality*, Humanities Press, New York 1963.
- Shipley, Thorne, *Intersensory Origin of Mind (A Revisit to Emergent Evolution)* (Routledge, Londra, 1995).
- Siegel, H. *Educating Reason; Rationality, Critical Thinking and Education* (Routledge, Londra 1992).
- Smith, J.C. (ed.), *Historical Foundations of Cognitive Science*, (Kluwer Acad. Publ. Boston, 1990).
- Smith, P., O. R. Jones, *The Philosophy of Mind; An Introduction* (New York, NY: Cambridge University Press, 1986).
- Smith, E. E. and D. L. Medin, *Categories and Concepts* (Cambridge, MA: Harvard University Press, 1981).
- Sober, E. „Panglossian Functionalism and the Philosophy of Mind“, *Synthese*, 64, 1985.
- Sorell, T., *Descartes* (Oxford: University Press, 1987).
- Sosa, E. (ed.), *Causation and Conditionals* (Oxford: University Press, 1975; Oxford Readings series).
- Sousa, Ronald de, *The Rationality of Emotion*, (Mitt Press, Cambridge SUA, 1987).
- Sperry, R.W. Mind-Brain Interaction, Mentalism and Dualism (în *Neurosciences* nr. 5, 1980).
- Sperry R. W. și H. P. Stapp, Structure and Significance of the „Consciousness Revolution“ în *Mind* 3/1987.

- Stich, Stephen P. *From Folk Psychology to Cognitive Science: The Case against Belief* (Cambridge, MA: The MIT Press/Bradford 1983).
- Stich, St. și Ted Warfield (ed), *Mental Representation. A Reader* (Blackwell, Oxford, 1994).
- Strawson, P.F., *Individuals. An Essay in Descriptive Metaphysics*, Methuen, Londra, 1959).
- Stroll, A., *Epistemology: New Essays in the Theory of Knowledge* (Harper et Row, New York, 1967).
- Swinburne, R., *Faith and Reason* (Oxford: University Press, 1981)
- Teichman, J., *The Mind and the Soul. An Introduction to the Philosophy of Mind* (Londra: Routledge, 1974).
- Thagard, Paul, *Computational Philosophy of Science* (Mit Press, Cambridge, SUA, 1988).
- Toulmin, S. – „Concept-Formation“, în *Philosophy and Psychology* nr. 12/1961.
- Trigg, R., *Pain and Emotion*, Oxford Univ Press, Oxford, 1970.
- Tuomela, R. (ed.), *Disposition* (D. Reidel, Dordrecht, 1987).
- Turing, A. M., „Computing Machinery and Intelligence“ în *Mind and Machines*, ed. A. R. Anderson, (Englewood Cliffs, NY, Prentice Hall, 1981).
- Tzonis, Alexander, *Hermes and the Golden Thinking Machine*, (Mitt Press, Londra, 1990).
- Varela, Tr. J., E. Thompson, E. Rosch, *The Embodied Mind (Cognitive Science and Human Experience)* (Mit Press, Cambridge, SUA, 1991).
- Vendler, Zeno, *The Matter of Minds*, (Clarendon Press, Oxford, 1984).
- Vesey, G.N.A. (ed.), *The Human Agent*, Royal Institute of Philosophy Lecture Series, vol. 1 (Londra: Macmillan, 1968; Brighton: Harvester).
- Vesey, G.N.A. (ed.) *Knowledge and Necessity*, Royal Institute of Philosophy Lecture Series, vol. 3 (Londra: Macmillan, 1970; Brighton: Harvester).
- Vesey, G.N.A. (ed), *Nature and Conduct*, Royal Institute of Philosophy Lecture Series, vol. 8 (Londra: Macmillan, 1976).
- Watson, R.A. *Representational Ideas. From Plato to Patricia Churchland*, (Kluwer Acad. Publ., Dordrecht, 1995).
- White, Stephen L., *The Unity of the Self*, (Mitt Press, Cambridge-SUA, 1991).
- Wilkes, K.V., Realism and Antirealism in Psychology, în *British Journal for the Philosophy of Science*, vol. 35, 1983.
- Wilkes, K.V., *Real People, Personal Identity without Thought Experiments* (Clarendon Press, Oxford, 1988).
- Wisdom, J. – *Paradox and Discovery*, (B. Blackwell, Oxford, 1965).
- Wittgenstein, L., *Lectures and Conversations on Aesthetics, Psychology & Religious Belief*, ed. C. Barrett (Oxford: Blackwell, 1966).
- Wittgenstein, L., *Tractatus Logico-Philosophicus*, rev. edn trad. de D.F. Pears and B.F. McGuinness, introd. B. Russell (Londra Routledge, 1961).
- Wittgenstein, L., *Philosophical Investigations*, trad. G.E.M. Amscombe (Oxford: Blackwell, 1967).
- Young, J.Z., *Philosophy and the Brain* (Oxford: University Press 1988).

# INDICE DE NOME

## A

Apel, K.O. – 141, 142  
 Alexander, R.O. – 34–35  
 Aristotel – 8, 119, 239, 240, 241  
 Armstrong, D.M.G. – 8, 92, 96, 239, 240, 241  
 Austin, J.L. – 203, 240  
 Aquinas – 15, 117

## B

Barnes, B. – 17  
 Benacerraf, P. – 230, 231  
 Berkeley – 19, 210  
 Bergson – 8, 283  
 Bergstrom, L. – 39  
 Bernard, Cl. – 274, 275  
 Bertalanffy, L. von, – 191  
 Besnier J.M. 308, 309  
 Blackburn, S. – 9  
 Blaga, L. – 275, 286  
 Block, N. – 9, 15, 24  
 Bloor, D. – 17  
 Bohm, D. – 278  
 Bohr, N. – 276, 283  
 Boltman, L. – 248, 249  
 Bradshaw, GF – 193  
 Braithwaite, – 240  
 Brand, M. – 13  
 Brentano, F. – 8, 50, 212, 213  
 Brown, J. – 39  
 Bunge, M. – 35, 191  
 Bühler, K. – 202, 203

## C

Campbell, D. – 256, 257  
 Carnap, R. – 124, 190  
 Capra, F. – 284  
 Chaitin, G. – 293, 299, 302, 307  
 Chisholm, R. – 123, 125  
 Chomsky, N. – 8  
 Church, D. – 15  
 Churchland, P.M. – 9, 10, 11, 15  
 Clifford, W.K. – 253, 254

Collingwood, R. G. – 189  
 Comte, A. – 189  
 Crane, T. – 9, 13, 35, 37, 38, 39, 89  
 Cummins, R.C. – 13, 32

## D

Davidson, D. – 9, 15, 24–25, 36, 39, 93, 94, 95, 96, 102–105, 142–149, 150–152, 154, 155, 156, 157, 158.  
 Darwin – 205, 206, 248, 252, 254, 255, 256, 259.  
 Dawkins, R. – 33, 219  
 Dennett, D.C. – 9, 13, 26, 72  
 Descartes – 8, 15, 121, 188, 210, 211, 241  
 Dewey, J. – 276, 287  
 Dilthey, W. – 142, 189  
 Dray H. – 143, 154, 155  
 Dretske, F. – 76, 110, 239  
 Durkheim, E. – 189, 191

## E

Eccles, J. – 15, 117  
 Edmonds, B. – 302, 303  
 Eliade, M. – 29, 30  
 Einstein, A. – 229, 231, 235  
 Emerson – 20  
 Evans, G. – 61, 63  
 Espagnat, B. – 277, 281, 283

## F

Feigl, H. – 15, 16  
 Fisher, R. A. – 246–247  
 Fodor, JA. – 9, 12, 13, 15, 24, 62, 76, 106, 316  
 Follesdal, D. – 150  
 Foucault, M. – 29, 30  
 Frege G. – 201, 202  
 Freud, S. – 9, 15, 39, 211, 212, 278, 283, 288

## G

Galois, E – 277  
Garfield, J.L. – 12, 13  
Gadamer, S. – 72, 73  
Glymour, C – 76, 78  
Guilford, J. P. – 288  
Galilei – 231, 238, 239, 246, 247  
Gödel, K – 251, 279, 294

## H

Habermas, J. – 8, 28, 30  
Hacker, PMS – 9, 12, 20, 39  
Hacking, J – 17, 18  
Haldane, J. J. – 38, 117  
Harré, R. – 9, 17  
Hempel, C.G. – 39, 124, 242, 149, 152,  
153, 154, 155, 158  
Honderich, T – 9, 12, 36, 37, 38, 39, 41  
Hume – 8, 19, 27, 36, 150, 188, 210, 239  
Husserl – 8, 33

## J

Jackson, F. – 110  
James W. – 8, 20  
Jung – 8, 277, 283  
Jaspers, K – 275

## K

Kant, I – 8, 18, 33, 36, 200  
Kepler – 193  
Kim, J. – 93, 98, 143, 157  
Knorr-Cetina – 17, 18

## L

Langley, P. W. – 193  
Latour, V.F. – 17  
Leibniz – 8, 218, 238, 241

Lewis, H.D. – 37, 92, 94, 95, 96, 112,  
257  
Locke – 20, 210, 211  
Lorenz, K – 33, 255  
Loschmidt, I – 248  
Lumsden, C.J. – 33, 34  
Lupasco, St – 279  
Lycan, W.G. – 37, 42  
Lynch, H. – 17  
Lyotard, E. – 30

## M

MacDonald, G – 98, 100, 101  
Mach – 254, 255  
MacIntyre, A.C. – 9, 13  
Malcom, N. – 39  
Maxwell – 235, 250  
McCulloch, W.S. – 12  
McDonald, C. – 98  
McFetridge, I – 117  
McGinn, C – 37  
McLanghlin, B. – 158–159  
Mead, GN – 184  
Mellor, DH. – 9, 12  
Merleau-Ponty, M – 8, 175  
Milikan, R. – 17  
Mill, J. St – 36, 141–142, 181

## N

Nagel, Th – 15, 33, 226  
Newell, A – 191, 192  
Newton – 236  
Noica, C – 282, 286  
Nozick, V.R. – 9, 17

## O

Occam, – 8, 92, 96  
Osherson, D.N. – 17, 19  
Ornstein, RE – 13



## P

Papineau, D-8, 36, 37, 39, 58, 76, 93,  
94, 95, 96  
Parfit, D - 9, 25  
Pascal, B - 280  
Peacocke, C - 13, 37, 93  
Platon - 8, 15, 241  
Place, U.T. - 91  
Pickering, A - 17  
Pettit, Ph-97, 100, 110  
Pollock, J - 97, 100, 110  
Peirce - 8, 210  
Podolsky-Rosen - 237-38  
Popper, K.R. -15, 30, 38, 177, 197  
Putnam, H-13, 15, 24, 26, 61, 97  
Pylyshin, L - 13, 316

## Q

Quine, W.O. - 15, 25, 32, 157, 159

## R

Ramsey - 8  
Reid, F-8, 19  
Ricoeur, P. - 29  
Rorty, R-9, 18, 31, 32  
Russell, B-36, 201, 202  
Ryle, W. - 9, 22, 23

## S

Schrodinger M. - 236  
Searle, J.R. - 9, 25, 27, 36, 43-50, 52, 53  
Sellars, W. - 9, 23-24, 32, 123

Simon, H.A. - 191, 192, 309  
Skinner, B.F. - 12  
Smart, H., - 15, 91, 92  
Smolensky, P. - 10  
Sorensen, R.-39, 243  
Sober, E - 264  
Sperry, RW - 9, 281  
Stalnaker, R. -76, 128, 132, 258  
Stich, S.P. - 10, 12, 260

## T

Tarski - 15  
Teilhard de Chardin - 8, 291  
Tertulian - 175, 176  
Titchener S. - 8  
Tooley, M - 240, 241, 244  
Turing, AM - 18

## W

Weber, M. - 199  
Whewell, W. - 251, 253  
Wiggins - 15  
Wilkes, K.V. - 38, 39  
Williams, C. - 267  
Wilson, F. - 18, 33  
Wittgenstein, L. - 8, 12, 18, 20, 21, 22,  
32, 117  
Wright - 76, 142  
Wundt - 8, 277

## Y

Yablo, St. - 110, 112

colecția

# BIBLIOTECA DE ȘTIINȚE

seria • Științe umaniste

## Filosofia mentalului

O nouă disciplină filosofică ?

Ce înseamnă a gândi, a simți, a-ți aminti, a experimenta ? Formează ele o unitate ?

Răspund la aceste întrebări: K.R. Popper, T. Honderich, D. Papineau, Tim Crane, J.J. Haldane, K.V. Wilkes, G.J. Lock (Anglia), R. Sorensen (S.U.A.), J.R. Brown (Canada), L. Bergstrom (Suedia), și A. Botez, Mircea și Mihaela Malița, L. Miroș, I. Mânzat, G. Ștefan și G.G. Costandache (România).